```
OM of: US-09-674-779-2 to: GenEmbl:*   out_format : pfs

Date: Sep 16, 2002  7:55 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124143_26862/app_query.fasta_1.310
 -DB=GenEmbl -QFMT=fastap -SUFFIX=oli.rge -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09674779_@CGN1_1_4034 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 1844.010000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pat:AX081145         +  250.00 5248.11  5.8e-284  753   ! AX081145 Sequence 1 from Patent
gb_pat:AX067442         +  250.00 5221.80  1.7e-282  23210  ! AX067442 Sequence 17 from Pat
gb_htg:AC097141         -    9.00  120.01    254.01  157901  ! AC097141 Rattus norvegicus c
gb_pr:AC068544          -    9.00  119.94    256.38  159397  ! AC068544 Homo sapiens BAC cl
gb_htg:AC106164         -    9.00  119.05    287.28  178928  ! AC106164 Rattus norvegicus c
gb_htg:AC026821         +    9.00  118.74    299.06  186385  ! AC026821 Homo sapiens chromo
gb_pr:AL355499          -    9.00  118.27    317.62  198141  ! AL355499 Human DNA sequence
gb_htg:AL512373         -    9.00  117.83    335.99  209790  ! AL512373 Homo sapiens chromo
gb_pat:AX081147         +    8.00  162.46      1.10      40  ! AX081147 Sequence 3 from Patent
gb_pat:AX081148         -    8.00  159.35      1.64      60  ! AX081148 Sequence 4 from Patent
gb_sts:G22057           -    8.00  144.23     11.37     430  ! G22057 human STS WI-13488, sequ
gb_sts:G61482           +    8.00  141.06     17.07     650  ! G61482 SHGC-85859 Human Homo sa
em_htgo_inv:AC085197 +    8.00  138.96     22.33     854  ! Ac085197 Giardia intestinalis c
gb_ba:MSGLIPOPRE        -    8.00  138.56     23.52     900  ! L08229 Mycobacterium intracellu
em_htgo_inv:AC028827 +    8.00  138.24     24.52     939  ! Ac028827 Giardia intestinalis c
em_htgo_inv:AC071017 -    8.00  137.96     25.42     974  ! Ac071017 Giardia intestinalis c
em_htgo_inv:AC060939 +    8.00  137.88     25.68     984  ! Ac060939 Giardia intestinalis c
em_htgo_inv:AC076436 +    8.00  137.74     26.14    1002  ! Ac076436 Giardia intestinalis
gb_pl:ATH271473         -    8.00  136.19     31.88    1226  ! AJ271473 Arabidopsis thaliana
gb_pat:E03308           -    8.00  135.97     32.80    1262  ! E03308 DNA encoding recombinan
gb_pl:AF370181          -    8.00  135.13     36.51    1407  ! AF370181 Arabidopsis thaliana
gb_ov:IPJ00267          +    8.00  132.94     48.36    1872  ! AJ000267 Ictalurus punctatus m
gb_pat:AR156086         -    8.00  132.41     51.79    2007  ! AR156086 Sequence 7 from paten
gb_vi:FCCPPKS19         -    8.00  132.39     51.87    2010  ! X99446 Feline calicivirus gene
gb_pat:E12464           -    8.00  132.30     52.48    2034  ! E12464 Capsid protein gene of
gb_pr:AK022355          -    8.00  132.14     53.57    2077  ! AK022355 Homo sapiens cDNA FLJ
gb_ba:AB010463          +    8.00  131.55     57.78    2243  ! AB010463 Vibrio parahaemolytic
gb_vi:CVXCAPS           -    8.00  130.74     64.12    2493  ! M32819 Feline calicivirus caps
gb_ov:AF072657          +    8.00  129.51     75.09    2927  ! AF072657 Danio rerio HLH-conta
gb_pr:AY032952          -    8.00  128.14     89.54    3500  ! AY032952 Homo sapiens nGAP-lik
gb_vi:FCL3E             -    8.00  128.10     89.95    3516  ! D90357 Feline calicivirus (FCV
gb_pr:AB051530          -    8.00  127.53     96.84    3790  ! AB051530 Homo sapiens mRNA for
gb_ro:AF236130          -    8.00  123.84    155.42    6128  ! AF236130 Rattus norvegicus DOC
gb_pat:AX251205         -    8.00  123.64    159.48    6291  ! AX251205 Sequence 173 from Pat
gb_vi:FCU13992          -    8.00  122.11    194.02    7677  ! U13992 Feline calicivirus CFI/
gb_vi:FCLF4             -    8.00  122.10    194.12    7681  ! D31836 Feline calicivirus genc
gb_ba:AF400582          +    8.00  121.45    211.17    8367  ! AF400582 Acinetobacter sp. ADP
gb_ba:AE002287          +    8.00  119.50    271.00   10780  ! AE002287 Chlamydia muridarum,
gb_ba:AE003962          -    8.00  119.36    276.10   10986  ! AE003962 Xylella fastidios
gb_ba:AE001354          +    8.00  118.83    295.44   11768  ! AE001354 Chlamydia trachom
gb_htg:AC110326         -    8.00  115.28    466.00   18696  ! AC110326 Rattus norvegicus
gb_in:CEK07A1           -    8.00  109.62    963.09   39086  ! Z81097 Caenorhabditis eleg
gb_pr:AL357372          -    8.00  108.79   1.1e+03   43553  ! AL357372 Human DNA sequenc
gb_htg:AC098224         -    8.00  108.63   1.1e+03   44428  ! AC098224 Rattus norvegicus
gb_htg:AC014969         -    8.00  108.32   1.1e+03   46275  ! AC014969 Drosophila melano

seq_name: gb_pat:AX081145

seq_documentation_block:
LOCUS       AX081145               753 bp    DNA     linear   PAT 27-FEB-2001
DEFINITION  Sequence 1 from Patent WO0109335.
ACCESSION   AX081145
VERSION     AX081145.1  GI:13170040
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Thonnard,J.
  TITLE     Moraxella catarrhalis antigen basb120
  JOURNAL   Patent: WO 0109335-A 1 08-FEB-2001;
            SmithKline Beecham Biologicals s.a. (BE)
FEATURES             Location/Qualifiers
     source          1. .753
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
BASE COUNT      239 a    166 c    152 g    196 t
ORIGIN


alignment_scores:
           Quality:  250.00              Length:    250
             Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AX081145   ..

 Align seg 1/1  to: AX081145  from: 1  to: 753

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 50

      17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 100

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 150

      51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 200

      67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 250

      84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 300

     101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 350

     117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    351 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 400

    134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 450

    151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 500

    167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 550

    184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    551 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 600

    201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    601 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 650

    217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    651 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 700

    234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    701 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 750
```

```
seq_name: gb_pat:AX067442

seq_documentation_block:
LOCUS       AX067442              23210 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 17 from Patent WO0078968.
ACCESSION   AX067442
VERSION     AX067442.1  GI:12545062
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 23210)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 17 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .23210
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
     unsure          28
                     /note="a or g or c or t, unknown, or other"
BASE COUNT     7108 a   4482 c   4951 g   6668 t      1 others
ORIGIN


alignment_scores:
            Quality:  250.00             Length:    250
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AX067442   ..

 Align seg 1/1  to: AX067442  from: 1  to: 23210

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11483 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 11532

      17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
   11533 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 11582

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11583 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 11632

      51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11633 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 11682

      67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11683 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 11732

      84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11733 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 11782

     101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11783 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 11832

     117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11833 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 11882

     134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11883 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 11932

     151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11933 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 11982

     167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11983 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 12032

     184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   12033 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 12082

     201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   12083 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 12132

     217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   12133 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 12182

     234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   12183 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 12232
```

```
seq_name: gb_htg:AC097141

seq_documentation_block:
LOCUS       AC097141              157901 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-181K9, *** SEQUENCING IN PROGRESS
            ***, 63 unordered pieces.
ACCESSION   AC097141
VERSION     AC097141.3  GI:17973578
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 157901)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
```

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157901)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Dec 20, 2001 this sequence version replaced gi:17064462.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GICJ
Center clone name: CH230-181K9
---------- Summary Statistics
Assembly program: Phrap; version 0.990329First call to findPhrapList
Consensus quality: 122789 bases at least Q40
Consensus quality: 133264 bases at least Q30
Consensus quality: 142961 bases at least Q20
Estimated insert size: 131949; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
----------
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*        1     4480: contig of 4480 bp in length
*     4481     4580: gap of unknown length
*     4581     9598: contig of 5018 bp in length
*     9599     9698: gap of unknown length
*     9699    15571: contig of 5873 bp in length
*    15572    15671: gap of unknown length
*    15672    20138: contig of 4467 bp in length
*    20139    20238: gap of unknown length
*    20239    24260: contig of 4022 bp in length
*    24261    24360: gap of unknown length
*    24361    28149: contig of 3789 bp in length
*    28150    28249: gap of unknown length
*    28250    32736: contig of 4487 bp in length
*    32737    32836: gap of unknown length
*    32837    35463: contig of 2627 bp in length
*    35464    35563: gap of unknown length
*    35564    40210: contig of 4647 bp in length
*    40211    40310: gap of unknown length
*    40311    43598: contig of 3288 bp in length
*    43599    43698: gap of unknown length
*    43699    45945: contig of 2247 bp in length
*    45946    46045: gap of unknown length
*    46046    49825: contig of 3780 bp in length
*    49826    49925: gap of unknown length
*    49926    54719: contig of 4794 bp in length
*    54720    54819: gap of unknown length
*    54820    58716: contig of 3897 bp in length
*    58717    58816: gap of unknown length
*    58817    63070: contig of 4254 bp in length
*    63071    63170: gap of unknown length
*    63171    66484: contig of 3314 bp in length
*    66485    66584: gap of unknown length
*    66585    69477: contig of 2893 bp in length
*    69478    69577: gap of unknown length
*    69578    71444: contig of 1867 bp in length
*    71445    71544: gap of unknown length
*    71545    74171: contig of 2627 bp in length
*    74172    74271: gap of unknown length
*    74272    76780: contig of 2509 bp in length
*    76781    76880: gap of unknown length
*    76881    79690: contig of 2810 bp in length
*    79691    79790: gap of unknown length
*    79791    82447: contig of 2657 bp in length
*    82448    82547: gap of unknown length
*    82548    84890: contig of 2343 bp in length
*    84891    84990: gap of unknown length
*    84991    88153: contig of 3163 bp in length
*    88154    88253: gap of unknown length
*    88254    90755: contig of 2502 bp in length
*    90756    90855: gap of unknown length
*    90856    93299: contig of 2444 bp in length
*    93300    93399: gap of unknown length
*    93400    96121: contig of 2722 bp in length
*    96122    96221: gap of unknown length
*    96222    98033: contig of 1812 bp in length
*    98034    98133: gap of unknown length
*    98134   100571: contig of 2438 bp in length
*   100572   100671: gap of unknown length
*   100672   103441: contig of 2770 bp in length
*   103442   103541: gap of unknown length
*   103542   105524: contig of 1983 bp in length
*   105525   105624: gap of unknown length
*   105625   107507: contig of 1883 bp in length
*   107508   107607: gap of unknown length
*   107608   109471: contig of 1864 bp in length
*   109472   109571: gap of unknown length
*   109572   111494: contig of 1923 bp in length
*   111495   111594: gap of unknown length
*   111595   113470: contig of 1876 bp in length
*   113471   113570: gap of unknown length
*   113571   116166: contig of 2596 bp in length
*   116167   116266: gap of unknown length
*   116267   118544: contig of 2278 bp in length
*   118545   118644: gap of unknown length
```

```
          *   118645   120674: contig of 2030 bp in length
          *   120675   120774: gap of unknown length
          *   120775   122285: contig of 1511 bp in length
          *   122286   122385: gap of unknown length
          *   122386   123737: contig of 1352 bp in length
          *   123738   123837: gap of unknown length
          *   123838   125540: contig of 1703 bp in length
          *   125541   125640: gap of unknown length
          *   125641   126909: contig of 1269 bp in length
          *   126910   127009: gap of unknown length
          *   127010   128225: contig of 1216 bp in length
          *   128226   128325: gap of unknown length
          *   128326   129438: contig of 1113 bp in length
          *   129439   129538: gap of unknown length
          *   129539   131623: contig of 2085 bp in length
          *   131624   131723: gap of unknown length
          *   131724   133593: contig of 1870 bp in length
          *   133594   133693: gap of unknown length
          *   133694   135261: contig of 1568 bp in length
          *   135262   135361: gap of unknown length
          *   135362   136751: contig of 1390 bp in length
          *   136752   136851: gap of unknown length
          *   136852   137975: contig of 1124 bp in length
          *   137976   138075: gap of unknown length
          *   138076   139657: contig of 1582 bp in length
          *   139658   139757: gap of unknown length
          *   139758   140811: contig of 1054 bp in length
          *   140812   140911: gap of unknown length
          *   140912   142409: contig of 1498 bp in length
          *   142410   142509: gap of unknown length
          *   142510   144063: contig of 1554 bp in length
          *   144064   144163: gap of unknown length
          *   144164   145707: contig of 1544 bp in length
          *   145708   145807: gap of unknown length
          *   145808   146876: contig of 1069 bp in length
          *   146877   146976: gap of unknown length
          *   146977   148337: contig of 1361 bp in length
          *   148338   148437: gap of unknown length

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC097141/rev   ..

 Align seg 1/1  to reverse of: AC097141  from: 1  to: 157901

       6 GlnTyrPheIleThrThrLeuIleSer 14
         ||||||||||||||||||||||||||||
   52516 CAGTATTTTATAACGACCTTGATATCT 52490

seq_name: gb_pr:AC068544

seq_documentation_block:
LOCUS       AC068544              159397 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-369J9 from 2, complete sequence.
ACCESSION   AC068544
VERSION     AC068544.7  GI:15920133
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159397)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 159397)
  AUTHORS   Harris,A., Cordum,H., Dignan,G. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-369J9
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 159397)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 159397)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 159397)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 4, 2001 this sequence version replaced gi:14333968.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0369J09
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-308G19; the clone sequenced
            to the right is RP11-84A12.  Actual start of this clone is at base
            position 1 of RP11-369J9; actual end is at base position 159397 of
            RP11-369J9.

            Data from AC012486 was used to finish this clone, AC068544.
            Polymorphisms have been identified between AC068544 and AC012486.

            Unresolved dinucleotide repeat from base position 14017 to 14486.
FEATURES                Location/Qualifiers
     source             1. .159397
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
```

```
                     /clone="RP11-369J9"
                     /clone_lib="RPCI-11"
     repeat_region   112. .227
                     /rpt_family="L2"
     repeat_region   289. .380
                     /rpt_family="CR1"
     repeat_region   503. .523
                     /rpt_family="AT_rich"
     repeat_region   570. .671
                     /rpt_family="CR1"
     repeat_region   1108. .1263
                     /rpt_family="L2"
     repeat_region   1584. .1648
                     /rpt_family="GA-rich"
     repeat_region   3550. .3605
                     /rpt_family="AT_rich"
     repeat_region   3740. .4235
                     /rpt_family="L2"
     repeat_region   4223. .4255
                     /rpt_family="(TGAA)n"
     repeat_region   4277. .4368
                     /rpt_family="L2"
     repeat_region   4693. .4822
                     /rpt_family="AcHobo"
     repeat_region   4842. .5142
                     /rpt_family="Alu"
     repeat_region   5124. .5143
                     /rpt_family="(TAAA)n"
     repeat_region   6061. .6192
                     /rpt_family="AcHobo"
     misc_feature    8244. .8813
                     /note="similar to EST BE985413 (NID:g10658696)"
     misc_feature    8591. .9720
                     /note="CpG_island (%GC=75.5, o/e=0.90, #CpGs=161)"
     repeat_region   8733. .8760
                     /rpt_family="GC_rich"
     repeat_region   8766. .8800
                     /rpt_family="GC_rich"
     repeat_region   8825. .8848
                     /rpt_family="GC_rich"
     repeat_region   9243. .9289
                     /rpt_family="GC_rich"
     repeat_region   10181. .10660
                     /rpt_family="MaLR"
     repeat_region   10896. .10994
                     /rpt_family="Alu"
     repeat_region   10995. .11281
                     /rpt_family="L1"
     repeat_region   11268. .11296
                     /rpt_family="(TAAAAA)n"
     repeat_region   11282. .11474
                     /rpt_family="Alu"
     repeat_region   11449. .11544
                     /rpt_family="GA-rich"
     repeat_region   11825. .12183
                     /rpt_family="CR1"
     repeat_region   12210. .12312
                     /rpt_family="AcHobo"
     repeat_region   12313. .12651
                     /rpt_family="MaLR"
     repeat_region   12652. .12797
                     /rpt_family="AcHobo"
     repeat_region   12854. .13010
                     /rpt_family="AcHobo"
     repeat_region   13012. .13390
                     /rpt_family="AcHobo"
     repeat_region   13356. .13387
                     /rpt_family="(TTG)n"
     repeat_region   13391. .13681
                     /rpt_family="Alu"
     repeat_region   13682. .13798
                     /rpt_family="AcHobo"
     repeat_region   13807. .14038
                     /rpt_family="Alu"
     repeat_region   14022. .14071
                     /rpt_family="GA-rich"
     repeat_region   14072. .14334
                     /rpt_family="G-rich"
     repeat_region   14308. .14478
                     /rpt_family="GA-rich"
     repeat_region   14490. .14534
                     /rpt_family="AcHobo"
     repeat_region   14581. .14869
                     /rpt_family="Alu"
     repeat_region   15548. .15695
                     /rpt_family="L2"
     repeat_region   15745. .15832
                     /rpt_family="MaLR"
     repeat_region   15925. .15970
                     /rpt_family="MaLR"
     repeat_region   16115. .16455
                     /rpt_family="MaLR"
     repeat_region   16567. .16660
                     /rpt_family="L2"
     repeat_region   16661. .16852
                     /rpt_family="Alu"
     repeat_region   16845. .16890
                     /rpt_family="(GGAGAA)n"
     repeat_region   17631. .17661
                     /rpt_family="(CTG)n"
     repeat_region   18253. .18341
                     /rpt_family="MIR"
     repeat_region   18799. .19077
                     /rpt_family="Alu"
     repeat_region   19077. .19189
                     /rpt_family="(TATATG)n"
     repeat_region   19669. .19797
                     /rpt_family="ERVL"
     repeat_region   19798. .20105
                     /rpt_family="Alu"

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC068544/rev   ..

 Align seg 1/1  to reverse of: AC068544  from: 1  to: 159397

      28 ProGlnValSerProIleLysThrPro 36
         |||||||||||||||||||||||||||
   15836 CCACAAGTGAGCCCCATCAAAACTCCT 15810

seq_name: gb_htg:AC106164

seq_documentation_block:
LOCUS       AC106164               178928 bp    DNA     linear   HTG 12-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-97I17, *** SEQUENCING IN PROGRESS
            ***, 83 unordered pieces.
ACCESSION   AC106164
VERSION     AC106164.1  GI:18138685
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 178928)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
```

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 178928)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GJKA
Center clone name: CH230-97I17
---------- Summary Statistics
Assembly program: Phrap; version 0.990329First call to findPhrapList
Consensus quality: 145158 bases at least Q40
Consensus quality: 158004 bases at least Q30
Consensus quality: 167877 bases at least Q20
Estimated insert size: 133921; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation
----------
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 83 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7669: contig of 7669 bp in length
* 7670 7769: gap of unknown length
* 7770 11596: contig of 3827 bp in length
* 11597 11696: gap of unknown length
* 11697 17638: contig of 5942 bp in length
* 17639 17738: gap of unknown length
* 17739 21941: contig of 4203 bp in length
* 21942 22041: gap of unknown length
* 22042 26403: contig of 4362 bp in length
* 26404 26503: gap of unknown length
* 26504 30870: contig of 4367 bp in length
* 30871 30970: gap of unknown length
* 30971 34286: contig of 3316 bp in length
* 34287 34386: gap of unknown length
* 34387 36949: contig of 2563 bp in length
* 36950 37049: gap of unknown length
* 37050 39678: contig of 2629 bp in length
* 39679 39778: gap of unknown length
* 39779 42848: contig of 3070 bp in length
* 42849 42948: gap of unknown length
* 42949 45210: contig of 2262 bp in length
* 45211 45310: gap of unknown length
* 45311 47249: contig of 1939 bp in length
* 47250 47349: gap of unknown length
* 47350 49561: contig of 2212 bp in length
* 49562 49661: gap of unknown length
* 49662 52181: contig of 2520 bp in length
* 52182 52281: gap of unknown length
* 52282 54889: contig of 2608 bp in length
* 54890 54989: gap of unknown length
* 54990 58620: contig of 3631 bp in length
* 58621 58720: gap of unknown length
* 58721 60718: contig of 1998 bp in length
* 60719 60818: gap of unknown length
* 60819 63863: contig of 3045 bp in length
* 63864 63963: gap of unknown length
* 63964 67392: contig of 3429 bp in length
* 67393 67492: gap of unknown length
* 67493 71177: contig of 3685 bp in length
* 71178 71277: gap of unknown length
* 71278 74025: contig of 2748 bp in length
* 74026 74125: gap of unknown length
* 74126 76212: contig of 2087 bp in length
* 76213 76312: gap of unknown length
* 76313 77888: contig of 1576 bp in length
* 77889 77988: gap of unknown length
* 77989 80122: contig of 2134 bp in length
* 80123 80222: gap of unknown length
* 80223 82466: contig of 2244 bp in length
* 82467 82566: gap of unknown length
* 82567 86483: contig of 3917 bp in length
* 86484 86583: gap of unknown length
* 86584 88641: contig of 2058 bp in length
* 88642 88741: gap of unknown length
* 88742 92630: contig of 3889 bp in length
* 92631 92730: gap of unknown length
* 92731 95504: contig of 2774 bp in length
* 95505 95604: gap of unknown length
* 95605 97703: contig of 2099 bp in length
* 97704 97803: gap of unknown length
* 97804 99593: contig of 1790 bp in length
* 99594 99693: gap of unknown length
* 99694 102118: contig of 2425 bp in length
* 102119 102218: gap of unknown length
* 102219 103623: contig of 1405 bp in length
* 103624 103723: gap of unknown length
* 103724 104942: contig of 1219 bp in length
* 104943 105042: gap of unknown length
* 105043 106760: contig of 1718 bp in length
* 106761 106860: gap of unknown length
* 106861 109011: contig of 2151 bp in length
* 109012 109111: gap of unknown length
* 109112 110470: contig of 1359 bp in length
* 110471 110570: gap of unknown length
* 110571 112304: contig of 1734 bp in length
* 112305 112404: gap of unknown length
* 112405 114135: contig of 1731 bp in length

```
            *   114136   114235: gap of unknown length
            *   114236   115531: contig of 1296 bp in length
            *   115532   115631: gap of unknown length
            *   115632   116679: contig of 1048 bp in length
            *   116680   116779: gap of unknown length
            *   116780   118262: contig of 1483 bp in length
            *   118263   118362: gap of unknown length
            *   118363   120357: contig of 1995 bp in length
            *   120358   120457: gap of unknown length
            *   120458   121770: contig of 1313 bp in length
            *   121771   121870: gap of unknown length
            *   121871   122872: contig of 1002 bp in length
            *   122873   122972: gap of unknown length
            *   122973   124417: contig of 1445 bp in length
            *   124418   124517: gap of unknown length
            *   124518   126214: contig of 1697 bp in length
            *   126215   126314: gap of unknown length
            *   126315   127700: contig of 1386 bp in length
            *   127701   127800: gap of unknown length
            *   127801   129440: contig of 1640 bp in length
            *   129441   129540: gap of unknown length
            *   129541   130686: contig of 1146 bp in length
            *   130687   130786: gap of unknown length
            *   130787   132374: contig of 1588 bp in length
            *   132375   132474: gap of unknown length
            *   132475   133790: contig of 1316 bp in length
            *   133791   133890: gap of unknown length
            *   133891   135752: contig of 1862 bp in length
            *   135753   135852: gap of unknown length
            *   135853   137250: contig of 1398 bp in length
            *   137251   137350: gap of unknown length
            *   137351   138999: contig of 1649 bp in length
            *   139000   139099: gap of unknown length
            *   139100   140877: contig of 1778 bp in length
            *   140878   140977: gap of unknown length
            *   140978   141979: contig of 1002 bp in length

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC106164/rev   ..

 Align seg 1/1  to reverse of: AC106164  from: 1  to: 178928

     89 SerArgLeuGlyAsnTyrLeuProPro 97
        ||||||||||||||||||||||||||
  12111 TCTAGGCTGGGGAATTACCTTCCTCCA 12085

seq_name: gb_htg:AC026821

seq_documentation_block:
LOCUS       AC026821              186385 bp    DNA     linear   HTG 19-APR-2000
DEFINITION  Homo sapiens chromosome 9 clone RP11-146H11 map 9, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC026821
VERSION     AC026821.2  GI:7596844
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186385)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 9, clone RP11-146H11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186385)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 19, 2000 this sequence version replaced gi:7321616.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7172
                Center clone name: 146_H_11
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 168457 bases at least Q40
                Consensus quality: 176745 bases at least Q30
                Consensus quality: 180379 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 182785; sum-of-contigs
                Quality coverage: 4.0 in Q20 bases; agarose-fp
                Quality coverage: 3.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1236: contig of 1236 bp in length
            *     1237  1336: gap of      100 bp
            *     1337     2427: contig of 1091 bp in length
            *     2428  2527: gap of      100 bp
            *     2528     3569: contig of 1042 bp in length
            *     3570  3669: gap of      100 bp
            *     3670     4862: contig of 1193 bp in length
            *     4863  4962: gap of      100 bp
            *     4963     6314: contig of 1352 bp in length
            *     6315  6414: gap of      100 bp
            *     6415     7809: contig of 1395 bp in length
            *     7810  7909: gap of      100 bp
            *     7910     9008: contig of 1099 bp in length
            *     9009  9108: gap of      100 bp
            *     9109    10239: contig of 1131 bp in length
            *    10240 10339: gap of      100 bp
            *    10340    11629: contig of 1290 bp in length
            *    11630 11729: gap of      100 bp
            *    11730    12780: contig of 1051 bp in length
            *    12781 12880: gap of      100 bp
            *    12881    14000: contig of 1120 bp in length
```

```
          *    14001 14100: gap of      100 bp
          *    14101    16937: contig of 2837 bp in length
          *    16938 17037: gap of      100 bp
          *    17038    18547: contig of 1510 bp in length
          *    18548 18647: gap of      100 bp
          *    18648    20611: contig of 1964 bp in length
          *    20612 20711: gap of      100 bp
          *    20712    22864: contig of 2153 bp in length
          *    22865 22964: gap of      100 bp
          *    22965    25088: contig of 2124 bp in length
          *    25089 25188: gap of      100 bp
          *    25189    27258: contig of 2070 bp in length
          *    27259 27358: gap of      100 bp
          *    27359    29599: contig of 2241 bp in length
          *    29600 29699: gap of      100 bp
          *    29700    31991: contig of 2292 bp in length
          *    31992 32091: gap of      100 bp
          *    32092    36046: contig of 3955 bp in length
          *    36047 36146: gap of      100 bp
          *    36147    39220: contig of 3074 bp in length
          *    39221 39320: gap of      100 bp
          *    39321    44708: contig of 5388 bp in length
          *    44709 44808: gap of      100 bp
          *    44809    49064: contig of 4256 bp in length
          *    49065 49164: gap of      100 bp
          *    49165    53182: contig of 4018 bp in length
          *    53183 53282: gap of      100 bp
          *    53283    56337: contig of 3055 bp in length
          *    56338 56437: gap of      100 bp
          *    56438    61607: contig of 5170 bp in length
          *    61608 61707: gap of      100 bp
          *    61708    67507: contig of 5800 bp in length
          *    67508 67607: gap of      100 bp
          *    67608    74182: contig of 6575 bp in length
          *    74183 74282: gap of      100 bp
          *    74283    80327: contig of 6045 bp in length
          *    80328 80427: gap of      100 bp
          *    80428    89523: contig of 9096 bp in length
          *    89524 89623: gap of      100 bp
          *    89624    96073: contig of 6450 bp in length
          *    96074 96173: gap of      100 bp
          *    96174   105987: contig of 9814 bp in length
          *   105988 106087: gap of      100 bp
          *   106088   115891: contig of 9804 bp in length
          *   115892 115991: gap of      100 bp
          *   115992   128365: contig of 12374 bp in length
          *   128366 128465: gap of      100 bp
          *   128466   142661: contig of 14196 bp in length
          *   142662 142761: gap of      100 bp
          *   142762   160565: contig of 17804 bp in length
          *   160566 160665: gap of      100 bp
          *   160666   186385: contig of 25720 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="9"
                     /clone="RP11-146H11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1236
                     /note="assembly_fragment"
     misc_feature    1337. .2427
                     /note="assembly_fragment"
     misc_feature    2528. .3569
                     /note="assembly_fragment"
     misc_feature    3670. .4862
                     /note="assembly_fragment"
     misc_feature    4963. .6314
                     /note="assembly_fragment"
     misc_feature    6415. .7809
                     /note="assembly_fragment"
     misc_feature    7910. .9008
                     /note="assembly_fragment"
     misc_feature    9109. .10239
                     /note="assembly_fragment"
     misc_feature    10340. .11629
                     /note="assembly_fragment"
     misc_feature    11730. .12780
                     /note="assembly_fragment"
     misc_feature    12881. .14000
                     /note="assembly_fragment"
     misc_feature    14101. .16937
                     /note="assembly_fragment"
     misc_feature    17038. .18547
                     /note="assembly_fragment"
     misc_feature    18648. .20611
                     /note="assembly_fragment"
     misc_feature    20712. .22864
                     /note="assembly_fragment"
     misc_feature    22965. .25088
                     /note="assembly_fragment"
     misc_feature    25189. .27258
                     /note="assembly_fragment"
     misc_feature    27359. .29599
                     /note="assembly_fragment"
     misc_feature    29700. .31991
                     /note="assembly_fragment"
     misc_feature    32092. .36046
                     /note="assembly_fragment"
     misc_feature    36147. .39220
                     /note="assembly_fragment"
     misc_feature    39321. .44708
                     /note="assembly_fragment"
     misc_feature    44809. .49064
                     /note="assembly_fragment"
     misc_feature    49165. .53182
                     /note="assembly_fragment"

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC026821   ..

 Align seg 1/1  to: AC026821  from: 1  to: 186385

      16 MetLeuValAlaCysSerAlaProIle 24
         ||||||||||||||||||||||||||||
  149370 ATGCTGGTTGCTTGCTCTGCGCCCATC 149396

seq_name: gb_pr:AL355499

seq_documentation_block:
LOCUS        AL355499              198141 bp    DNA     linear   PRI 30-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-328K6 on chromosome 6, complete
             sequence.
ACCESSION    AL355499
VERSION      AL355499.15  GI:10086115
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 198141)
  AUTHORS    Beasley,O.
  TITLE      Direct Submission
  JOURNAL    Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Sep 11, 2000 this sequence version replaced gi:10039827.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
```

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-328K6 is from the library RPCI-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-328K6 The true left end of clone RP11-203H2 is at 153730 in this sequence.

| FEATURES | Location/Qualifiers |
|---|---|
| source | 1. .198141 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /chromosome="6" |
| | /clone="RP11-328K6" |
| | /clone_lib="RPCI-11.2" |
| repeat_region | 53. .404 |
| | /note="MLT1A1 repeat: matches 1. .365 of consensus" |
| repeat_region | 547. .644 |
| | /note="TIGGER2 repeat: matches 2398. .2500 of consensus" |
| repeat_region | 1036. .1337 |
| | /note="AluSg repeat: matches 1. .303 of consensus" |
| repeat_region | 1971. .2264 |
| | /note="AluSx repeat: matches 1. .293 of consensus" |
| repeat_region | 2482. .2708 |
| | /note="AluSx repeat: matches 2. .225 of consensus" |
| repeat_region | 2847. .3275 |
| | /note="Tigger2a repeat: matches 1. .433 of consensus" |
| repeat_region | 3385. .3695 |
| | /note="AluSx repeat: matches 1. .309 of consensus" |
| repeat_region | 4027. .4135 |
| | /note="L2 repeat: matches 2563. .2685 of consensus" |
| repeat_region | 5269. .5327 |
| | /note="L2 repeat: matches 2364. .2419 of consensus" |
| repeat_region | 6287. .6583 |
| | /note="AluSg repeat: matches 1. .297 of consensus" |
| repeat_region | 6836. .6940 |
| | /note="L1ME3A repeat: matches 5996. .6103 of consensus" |
| repeat_region | 7103. .7244 |
| | /note="MIR repeat: matches 20. .192 of consensus" |
| repeat_region | 7359. .7629 |
| | /note="MER7A repeat: matches 5. .268 of consensus" |
| repeat_region | 7631. .7918 |
| | /note="AluSp repeat: matches 1. .288 of consensus" |
| repeat_region | 7919. .7958 |
| | /note="20 copies 2 mer ca 100% conserved" |
| repeat_region | 8561. .9002 |
| | /note="MER83 repeat: matches 1. .448 of consensus" |
| repeat_region | 9110. .9410 |
| | /note="AluSx repeat: matches 1. .301 of consensus" |
| repeat_region | 9411. .9448 |
| | /note="19 copies 2 mer ta 100% conserved" |
| repeat_region | 9604. .9689 |
| | /note="MIR repeat: matches 41. .126 of consensus" |
| repeat_region | 9988. .10197 |
| | /note="L2 repeat: matches 2472. .2671 of consensus" |
| misc_feature | 10438. .10865 |
| | /note="CpG island" |
| | /evidence=not_experimental |
| repeat_region | 12263. .13439 |
| | /note="L1MC5 repeat: matches 6437. .7589 of consensus" |
| repeat_region | 13548. .13696 |
| | /note="MER5B repeat: matches 2. .159 of consensus" |
| repeat_region | 14380. .14549 |
| | /note="L2 repeat: matches 2534. .2690 of consensus" |
| repeat_region | 14550. .14840 |
| | /note="AluSg1 repeat: matches 1. .288 of consensus" |
| repeat_region | 14841. .15039 |
| | /note="L2 repeat: matches 2305. .2534 of consensus" |
| repeat_region | 15098. .15448 |
| | /note="MLT2CB repeat: matches 3. .411 of consensus" |
| repeat_region | 15449. .15807 |
| | /note="THE1B repeat: matches 1. .364 of consensus" |
| repeat_region | 15808. .15914 |
| | /note="MLT2CB repeat: matches 362. .466 of consensus" |
| repeat_region | 15927. .16206 |
| | /note="AluSg repeat: matches 1. .278 of consensus" |
| repeat_region | 16208. .16494 |
| | /note="L2 repeat: matches 2129. .2419 of consensus" |
| repeat_region | 16587. .16715 |
| | /note="L2 repeat: matches 2575. .2709 of consensus" |
| repeat_region | 17070. .17215 |
| | /note="MIR repeat: matches 100. .252 of consensus" |
| repeat_region | 17925. .17969 |
| | /note="L2 repeat: matches 2705. .2750 of consensus" |
| repeat_region | 18038. .18356 |
| | /note="AluSx repeat: matches 1. .312 of consensus" |
| repeat_region | 18394. .18566 |
| | /note="FRAM repeat: matches -1. .171 of consensus" |
| repeat_region | 18570. .18625 |
| | /note="28 copies 2 mer aa 75% conserved" |
| repeat_region | 19223. .19310 |
| | /note="MLT1J repeat: matches 103. .193 of consensus" |
| repeat_region | 19877. .20039 |
| | /note="LTR33 repeat: matches 350. .519 of consensus" |
| repeat_region | 20403. .20721 |
| | /note="AluSx repeat: matches 1. .308 of consensus" |
| repeat_region | 21963. .22029 |
| | /note="L1M4 repeat: matches 5721. .5794 of consensus" |
| repeat_region | 22047. .22340 |
| | /note="AluY repeat: matches 1. .294 of consensus" |
| repeat_region | 22864. .23282 |
| | /note="L2 repeat: matches 2027. .2491 of consensus" |
| repeat_region | 23285. .23570 |
| | /note="MLT1J repeat: matches 211. .514 of consensus" |
| repeat_region | 23748. .24062 |
| | /note="MLT1J repeat: matches 36. .389 of consensus" |
| repeat_region | 24289. .24401 |
| | /note="MER30 repeat: matches 11. .130 of consensus" |
| repeat_region | 24402. .24712 |
| | /note="AluY repeat: matches 1. .307 of consensus" |
| repeat_region | 24713. .24814 |
| | /note="MER30 repeat: matches 130. .229 of consensus" |
| repeat_region | 25096. .25391 |
| | /note="AluSx repeat: matches 1. .294 of consensus" |
| repeat_region | 25554. .25849 |
| | /note="AluJb repeat: matches 1. .295 of consensus" |
| repeat_region | 26135. .26440 |
| | /note="AluSq repeat: matches 1. .312 of consensus" |
| repeat_region | 27172. .27466 |
| | /note="AluSx repeat: matches 1. .296 of consensus" |
| repeat_region | 27883. .27997 |
| | /note="L1MA9 repeat: matches 6171. .6302 of consensus" |
| repeat_region | 28577. .28610 |
| | /note="17 copies 2 mer tt 85% conserved" |
| repeat_region | 28778. .28863 |
| | /note="MLT1-INTERNAL repeat: matches 431. .525 of consensus" |
| repeat_region | 28864. .29250 |
| | /note="MSTA repeat: matches 1. .426 of consensus" |
| repeat_region | 29251. .29276 |
| | /note="MLT1-INTERNAL repeat: matches 404. .431 of |

```
                     consensus"
     repeat_region   29277. .29575
                     /note="AluJb repeat: matches 12. .308 of consensus"
     repeat_region   29576. .29620
                     /note="MLT1-INTERNAL repeat: matches 356. .404 of
                     consensus"
     repeat_region   29621. .29980
                     /note="MSTB repeat: matches 1. .357 of consensus"
     repeat_region   29981. .30274
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   30275. .30344
                     /note="MSTB repeat: matches 357. .426 of consensus"
     repeat_region   30345. .30498
                     /note="MLT1-INTERNAL repeat: matches 246. .356 of
                     consensus"
     repeat_region   30601. .30889
                     /note="AluSq repeat: matches 1. .293 of consensus"
     repeat_region   32002. .32290
                     /note="AluSp repeat: matches 1. .290 of consensus"
     repeat_region   32307. .32556
                     /note="L1ME3A repeat: matches 5894. .6163 of consensus"
     repeat_region   32575. .32979
                     /note="L1ME2 repeat: matches 5754. .6315 of consensus"
     repeat_region   32980. .33039
                     /note="L1ME3A repeat: matches 5849. .5908 of consensus"
     repeat_region   33707. .34022
                     /note="AluJo repeat: matches 2. .312 of consensus"
     repeat_region   34185. .34946
                     /note="L1PA16 repeat: matches 5368. .6145 of consensus"
     repeat_region   35004. .35121
                     /note="FLAM_A repeat: matches 6. .123 of consensus"
     repeat_region   35316. .35620
                     /note="AluJb repeat: matches 1. .306 of consensus"
     repeat_region   36193. .36494

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AL355499/rev   ..

 Align seg 1/1  to reverse of: AL355499  from: 1  to: 198141

      16 MetLeuValAlaCysSerAlaProIle 24
         ||||||||||||||||||||||||||||
   91692 ATGCTGGTTGCTTGCTCTGCGCCCATC 91666

seq_name: gb_htg:AL512373

seq_documentation_block:
LOCUS       AL512373              209790 bp    DNA     linear   HTG 12-SEP-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-564O14, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL512373
VERSION     AL512373.12  GI:15617280
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209790)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2001 this sequence version replaced gi:13273887.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA564O14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 209055 bases at least Q40
            Consensus quality: 209344 bases at least Q30
            Consensus quality: 209466 bases at least Q20
            Insert size: 209790; sum-of-contigs
            Insert size: 194715; 2.8% error; agarose-fp
            Quality coverage: 8.31x in Q20 bases; sum-of-contigs Quality
            coverage: 8.96x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .209790
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-564O14"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .209790
                     /note="assembly_fragment:04568"
BASE COUNT    58386 a  41949 c  43662 g  65793 t
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AL512373/rev   ..

 Align seg 1/1  to reverse of: AL512373  from: 1  to: 209790

      16 MetLeuValAlaCysSerAlaProIle 24
         ||||||||||||||||||||||||||||
   57293 ATGCTGGTTGCTTGCTCTGCGCCCATC 57267

seq_name: gb_pat:AX081147

seq_documentation_block:
LOCUS       AX081147                 40 bp    DNA     linear   PAT 27-FEB-2001
DEFINITION  Sequence 3 from Patent WO0109335.
ACCESSION   AX081147
VERSION     AX081147.1  GI:13170041
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Thonnard,J.
  TITLE     Moraxella catarrhalis antigen basb120
  JOURNAL   Patent: WO 0109335-A 3 08-FEB-2001;
            SmithKline Beecham Biologicals s.a. (BE)
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT       17 a       4 c       7 g      12 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
```

```
            Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AX081147   ..

 Align seg 1/1  to: AX081147  from: 1  to: 40

       1 MetLysAsnPheAsnGlnTyrPhe 8
         ||||||||||||||||||||||||
      16 ATGAAAAATTTTAATCAATACTTT 39

seq_name: gb_pat:AX081148

seq_documentation_block:
LOCUS        AX081148                 60 bp    DNA     linear   PAT 27-FEB-2001
DEFINITION   Sequence 4 from Patent WO0109335.
ACCESSION    AX081148
VERSION      AX081148.1  GI:13170042
KEYWORDS     .
SOURCE       synthetic construct.
  ORGANISM   synthetic construct
             artificial sequence.
REFERENCE    1  (bases 1 to 60)
  AUTHORS    Thonnard,J.
  TITLE      Moraxella catarrhalis antigen basb120
  JOURNAL    Patent: WO 0109335-A 4 08-FEB-2001;
             SmithKline Beecham Biologicals s.a. (BE)
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT       14 a      5 c     21 g     20 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AX081148/rev   ..

 Align seg 1/1  to reverse of: AX081148  from: 1  to: 60

     243 HisValLeuProLysAsnLysLeu 250
         |||||||||||||||||||||||
      60 CATGTCTTACCAAAAAATAAGCTA 37

seq_name: gb_sts:G22057

seq_documentation_block:
LOCUS        G22057                   430 bp    DNA     linear   STS 31-MAY-1996
DEFINITION   human STS WI-13488, sequence tagged site.
ACCESSION    G22057
VERSION      G22057.1  GI:1342383
KEYWORDS     STS; STS sequence; primer; sequence tagged site.
SOURCE       human STSs derived from sequences in dbEST and the Unigene
             collection.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 430)
  AUTHORS    Hudson,T.
  TITLE      Whitehead Institute/MIT Center for Genome Research; Physically
             Mapped STSs
  JOURNAL    Unpublished
COMMENT
             Contact: Thomas Hudson
             Whitehead Institute/MIT Center for Genome Research
             Whitehead Institute for Biomedical Research
             9 Cambridge Center, Cambridge MA 02142 USA
             Tel: 617 252 1900
             Fax: 617 252 1902
             Email: thudson@genome.wi.mit.edu

             Primer A: TTTTTTTTTACACATACCAAAAACA
             Primer B: AAGTCTTTGGATGATTTTTTTAAGG
             STS size: 135
             PCR Profile:
                   Presoak:
                   Denaturation:
                   Annealing: 56 degrees C
                   Polymerization:
                   PCR Cycles: 35
                   Thermal Cycler:
             Protocol:
                 Template: 10 ng
                 Primer: each 5 pM
                 dNTPs: each 4 nM
                 Taq Polymerase: 0.025 units/ul
                 Total Vol: 20 ul

             Buffer:
                 MgCl2: 1.5 mM
                 KCl: 50 mM
                 Tris-HCL: 10 mM
                 pH: 9.3

             Derived from dbEST (genbank accession R40672).
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="798.2 cR from top of Chr1 linkage group"
     STS             4. .138
     primer_bind     4. .28
     primer_bind     complement(114. .138)
BASE COUNT       115 a     81 c     84 g    146 t      4 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000              Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x G22057/rev   ..

 Align seg 1/1  to reverse of: G22057  from: 1  to: 430

      24 IleProThrAsnProGlnValSer 31
         |||||||||||||||||||||||
     374 ATACCCACNAATCCACAGGTGAGT 351

seq_name: gb_sts:G61482

seq_documentation_block:
LOCUS        G61482                   650 bp    DNA     linear   STS 30-MAR-2000
DEFINITION   SHGC-85859 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION    G61482
VERSION      G61482.1  GI:6126651
KEYWORDS     STS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 650)
  AUTHORS    Olivier,M. and Cox,D.R.
  TITLE      Unpublished, Olivier, M., Cox, D.R. (2000)
  JOURNAL    Unpublished
COMMENT
             Contact: Michael Olivier, David R. Cox
```

```
          Stanford Human Genome Center
          Stanford University School of Medicine
          4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
          Tel: (650) 320-5800
          Fax: (650) 320-5801
          Email: olivier@shgc.stanford.edu
          Primer A: TATCTCAGCAGTTTTGAGGGGAG
          Primer B: AGCTGAATCTGTCAACCCTGAAG
          STS size: 326
          PCR Profile:
                  Initial incubation: 95 degrees C for 10 minutes
                  Denaturation:       94 degrees C for 30 seconds
                  Annealing:          60 degrees C for 30 seconds
                  Polymerization:     72 degrees C for 23 seconds
                  PCR Cycles:         30
                  Thermal Cycler:     Perkin Elmer 9700
          Protocol:
                  Template:                   25 ng
                  Primer:                     each 1 uM
                  dNTPs:                      each 200 uM
                  AmpliTaq Gold Polymerase:   0.07 units/ul
                  Total Vol:                  5 ul

          Buffer:
                  MgCl2:          2.5 mM
                  KCl:            50 mM
                  Tris-HCl:       10 mM
                  pH:             8.3

          BAC ends sequenced at TIGR from the RPCI11 BAC library. Designed
          and developed at the Stanford Human Genome Center.
FEATURES            Location/Qualifiers
     source         1. .650
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="13"
                    /clone_lib="Human"
     STS            6. .331
     primer_bind    6. .28
     primer_bind    complement(309. .331)
BASE COUNT      185 a    109 c    144 g    212 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x G61482   ..

 Align seg 1/1  to: G61482  from: 1  to: 650

       8 PheIleThrThrLeuIleSerSer 15
         |||||||||||||||||||||||
     100 TTTATTACCACTTTAATTTCTTCA 123

seq_name: em_htgo_inv:AC085197

seq_documentation_block:
ID   AC085197   standard; DNA; HTG; 854 BP.
XX
AC   AC085197;
XX
SV   AC085197.1
XX
DT   02-DEC-2000 (Rel. 66, Created)
DT   02-DEC-2000 (Rel. 66, Last updated, Version 1)
XX
DE   Giardia intestinalis clone KJ4544 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-854
RX   MEDLINE; 20389616.
RA   McArthur A.G., Morrison H.G., Nixon J.E.J., Passamaneck N.Q.E., Kim U.,
RA   Hinkle G., Crocker M.K., Holder M.E., Farr R., Reich C.I., Olsen G.J.,
RA   Aley S.B., Adam R.D., Gillin F.D., Sogin M.L.;
RT   "The Giardia genome project database";
RL   FEMS Microbiol. Lett. 189(2):271-273(2000).
XX
RN   [2]
RP   1-854
RA   Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Fierro L.A.,
RA   Aley S.B., Sogin M.L.;
RT   ;
RL   Submitted (29-NOV-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *        1       854: contig of 854 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .854
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="KJ4544"
XX
SQ   Sequence 854 BP; 208 A; 200 C; 215 G; 228 T; 3 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC085197   ..

 Align seg 1/1  to: AC085197  from: 1  to: 854

      12 LeuIleSerSerMetLeuValAla 19
         ||||||||||||||||||||||||
     211 TTGATTTCTTCAATGTTGGTCGCA 234

seq_name: gb_ba:MSGLIPOPRE

seq_documentation_block:
LOCUS       MSGLIPOPRE              900 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Mycobacterium intracellulare lipoprotein precursor (Mi43) gene,
            complete cds.
ACCESSION   L08229
VERSION     L08229.1  GI:149971
KEYWORDS    lipoprotein; precursor protein.
SOURCE      Mycobacterium intracellulare (strain 13950) (library: lambda gt11)
            DNA.
  ORGANISM  Mycobacterium intracellulare
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
```

```
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium avium complex (MAC).
REFERENCE   1  (bases 1 to 900)
  AUTHORS   Nair,J., Rouse,D.A. and Morris,S.L.
  TITLE     Nucleotide sequence analysis and serologic characterization of a
            27-Kilodalton Mycobacterium intracellulare lipoprotein
  JOURNAL   Infect. Immun. 61, 1074-1081 (1993)
  MEDLINE   93162797
FEATURES             Location/Qualifiers
     source          1. .900
                     /organism="Mycobacterium intracellulare"
                     /strain="13950"
                     /db_xref="taxon:1767"
                     /tissue_lib="lambda gt11"
     RBS             13. .16
     CDS             25. .813
                     /standard_name="Mi43 lipoprotein precursor"
                     /function="unknown"
                     /note="lipoprotein concensus sequence is LSACGS in
                     precursor peptide."
                     /codon_start=1
                     /transl_table=11
                     /product="lipoprotein"
                     /protein_id="AAA25364.1"
                     /db_xref="GI:149972"
                     /translation="MSASCAVPRLTRFAVFAVAGATALSLSACGSSNKSSSTSTSTST
                     STSTVTSAAPSSTPNAEAKVSGLIASVAGNSIQVTKEDNATAAVNFTSATKITEAVPA
                     GLPDVTQGSCLIVKPTEGSAPGQPVTAAKVKISESVNGTCPKPHESTPGGASSTPPSG
                     SPSPAPAKPAWVRGSVASVSGDTINLTGTDASGNTTQTTVTVDDKTKYTKQTTANTEA
                     IAPGKCLSARGTTDSGGALQATSIKLRQAVDGKCGKPKQPGQGG"
     repeat_region   order(847. .858,868. .879)
                     /rpt_type=inverted
BASE COUNT      172 a    339 c    284 g    105 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x MSGLIPOPRE/rev   ..

 Align seg 1/1  to reverse of: MSGLIPOPRE  from: 1  to: 900

      16 MetLeuValAlaCysSerAlaPro 23
         ||||||||||||||||||||||||
     749 ATGCTCGTCGCCTGCAGCGCGCCG 726

seq_name: em_htgo_inv:AC028827

seq_documentation_block:
ID   AC028827   standard; DNA; HTG; 939 BP.
XX
AC   AC028827;
XX
SV   AC028827.1
XX
DT   05-APR-2000 (Rel. 63, Created)
DT   05-APR-2000 (Rel. 63, Last updated, Version 1)
XX
DE   Giardia intestinalis clone HF2271 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-939
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-939
RA   Hinkle G., Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U.,
RA   Crocker M.K., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (01-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *         1      939: contig of 939 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .939
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="HF2271"
XX
SQ   Sequence 939 BP; 212 A; 240 C; 218 G; 268 T; 1 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC028827   ..

 Align seg 1/1  to: AC028827  from: 1  to: 939

      12 LeuIleSerSerMetLeuValAla 19
         ||||||||||||||||||||||||
     889 TTGATTTCTTCAATGTTGGTCGCA 912

seq_name: em_htgo_inv:AC071017

seq_documentation_block:
ID   AC071017   standard; DNA; HTG; 974 BP.
XX
AC   AC071017;
XX
SV   AC071017.1
XX
DT   13-JUN-2000 (Rel. 64, Created)
DT   13-JUN-2000 (Rel. 64, Last updated, Version 1)
XX
DE   Giardia intestinalis clone EJ7174 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-974
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
```

```
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-974
RA   Eakin N.Q., Morrison H.G., McArthur A.G., Nixon J., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (06-JUN-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *        1       974: contig of 974 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .974
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="EJ7174"
XX
SQ   Sequence 974 BP; 271 A; 231 C; 241 G; 230 T; 1 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC071017/rev   ..

 Align seg 1/1  to reverse of: AC071017  from: 1  to: 974

      12 LeuIleSerSerMetLeuValAla 19
         ||||||||||||||||||||||||
     387 TTGATTTCTTCAATGTTGGTCGCA 364

seq_name: em_htgo_inv:AC060939

seq_documentation_block:
ID   AC060939   standard; DNA; HTG; 984 BP.
XX
AC   AC060939;
XX
SV   AC060939.1
XX
DT   24-APR-2000 (Rel. 63, Created)
DT   24-APR-2000 (Rel. 63, Last updated, Version 1)
XX
DE   Giardia intestinalis clone MJ4058 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-984
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
```

```
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-984
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (20-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *        1       984: contig of 984 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .984
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="MJ4058"
XX
SQ   Sequence 984 BP; 247 A; 234 C; 256 G; 247 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC060939   ..

 Align seg 1/1  to: AC060939  from: 1  to: 984

      12 LeuIleSerSerMetLeuValAla 19
         ||||||||||||||||||||||||
     264 TTGATTTCTTCAATGTTGGTCGCA 287

seq_name: em_htgo_inv:AC076436

seq_documentation_block:
ID   AC076436   standard; DNA; HTG; 1002 BP.
XX
AC   AC076436;
XX
SV   AC076436.1
XX
DT   02-AUG-2000 (Rel. 64, Created)
DT   02-AUG-2000 (Rel. 64, Last updated, Version 1)
XX
DE   Giardia intestinalis clone KJ4133 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-1002
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
```

```
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-1002
RA   Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (01-AUG-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *        1     1002: contig of 1002 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .1002
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="KJ4133"
XX
SQ   Sequence 1002 BP; 245 A; 237 C; 248 G; 271 T; 1 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC076436   ..

 Align seg 1/1  to: AC076436  from: 1  to: 1002

      12 LeuIleSerSerMetLeuValAla 19
         ||||||||||||||||||||||||
     403 TTGATTTCTTCAATGTTGGTCGCA 426

seq_name: gb_pl:ATH271473

seq_documentation_block:
LOCUS        ATH271473              1226 bp    mRNA    linear   PLN 25-JAN-2000
DEFINITION   Arabidopsis thaliana mRNA for hypothetical protein related to
             strictosidine synthase.
ACCESSION    AJ271473
VERSION      AJ271473.1  GI:6759490
KEYWORDS     strictosidine synthase.
SOURCE       thale cress.
  ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1  (bases 1 to 1226)
  AUTHORS    Mahon,P.
  JOURNAL    Thesis (2000) University of Cambridge
REFERENCE    2  (bases 1 to 1226)
  AUTHORS    Mahon,P.
  TITLE      Direct Submission
  JOURNAL    Submitted (18-JAN-2000) Mahon P., Biochemistry, Cambridge
             University, Tennis Court Road, Cambridge, CB2 1QW, UNITED KINGDOM
COMMENT      This sequence is from an unknown EST clone sent in error.

FEATURES             Location/Qualifiers
     source          1. .1226
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     CDS             <1. .1061
                     /function="related to strictosidine synthases"
                     /note="ORF"
                     /codon_start=3
                     /product="hypothetical protein"
                     /protein_id="CAB69786.1"
                     /db_xref="GI:6759491"
                     /translation="HASAIIGPDNLKGTKHVLQDAKTIPLPVDGPESLEFDPQGEGPY
                     VGVTDGRILKWRGEELGWVDFAYTSPHRDNCSSHEVVPSCGRPLGLSFERKTGDLYIC
                     DGYFGVMKVGPEGGLAELVVDEAEGRKVMFANQGDIDEEEDIFYFNDSSDTYHFRDVF
                     YVSLSGTKVGRVIRYDMKKKEAKVIMDKLRLPNGLALSKNGSFVVTCESSTNTCHRIW
                     VKGPKSGTNEVFATLPGSPDNIRRTPTGDFWVALHCKKNLFTRAVLIHTWVGRFFMNT
                     MKMETVIHFMNGGKPHGIVVKLSGETGEILEILEDSEGKTVKYVSEAYETKDGKLWIG
                     SVYWPAVWVLDTSVYDSI"
     polyA_site      1202
BASE COUNT      366 a    225 c    312 g    323 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ATH271473/rev   ..

 Align seg 1/1  to reverse of: ATH271473  from: 1  to: 1226

     144 ProAlaAsnThrGlnIleArgSer 151
         |||||||||||||||||||||||
    1020 CCGGCCAATACACAGATCCGATCC 997

seq_name: gb_pat:E03308

seq_documentation_block:
LOCUS        E03308                 1262 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION   DNA encoding recombinant streptokinase.
ACCESSION    E03308
VERSION      E03308.1  GI:2171525
KEYWORDS     JP 1992011892-A/1.
SOURCE       synthetic construct.
  ORGANISM   synthetic construct
             artificial sequence.
REFERENCE    1  (bases 1 to 1262)
  AUTHORS    Fujii,S., Katano,T., Majima,E., Ogino,K., Ono,K., Sakata,Y. and
             Uenoyama,T.
  TITLE      PROTEIN OF STREPTOKINASES, CORRESPONDING GENE, CORRESPONDING
             PLASMID RECOMBINANT, CORRESPONDING CHARACTER TRANSFORMANT AND
             PRODUCTION
  JOURNAL    Patent: JP 1992011892-A 1 16-JAN-1992;
             OTSUKA PHARMACEUT FACTORY INC
COMMENT      OS   Artificial gene
             OC   Artificial sequence; Genes.
             PN   JP 1992011892-A/1
             PD   16-JAN-1992
             PF   06-JUL-1990 JP 1990179851
             PR   11-JUL-1989 JP 89P   179432, 27-NOV-1989 JP 89P   307957, PR
             11-APR-1990 JP 90P    96830
             PI   FUJII SETSUO, KATANO TAMITAKA, MAJIMA EIJI, OGINO KOICHI, PI
             ONO KENJI,
             PI   SAKATA YASUYO, UENOYAMA TSUTOMU
             PC   C12N15/58,C12N1/21,C12N9/70,(C12N1/21,C12R1:19),(C12N9/70, PC
              C12R1:19);
             CC   strandedness: Double;
             CC   topology: Linear;
             CC   hypothetical: No;
             CC   anti-sense: No;
             CC   *source: clone=pSKX;
```

```
           FH   Key             Location/Qualifiers
           FH
           FT   5'UTR           1. .11
           FT   CDS             12. .1256
           FT                   /product='recombinant streptokinase' FT
           mat_peptide     12. .1253
           FT                   /product='recombinant streptokinase' FT
           3'UTR           1257. .1263.
FEATURES             Location/Qualifiers
     source          1. .1262
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT      365 a    350 c    262 g    285 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x E03308/rev   ..

 Align seg 1/1  to reverse of: E03308  from: 1  to: 1262

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
     393 GAGTTACAGAACCGTCTTTGTCAG 370

seq_name: gb_pl:AF370181

seq_documentation_block:
LOCUS       AF370181               1407 bp    mRNA    linear   PLN 30-APR-2001
DEFINITION  Arabidopsis thaliana unknown protein (F24I3.90/AT3g57010) mRNA,
            complete cds.
ACCESSION   AF370181
VERSION     AF370181.1  GI:13877836
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1407)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Yu,G.,
            Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
            Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
            Theologis,A.
  TITLE     Full Length cDNA of gene F24I3.90/AT3g57010 (GI:6911871)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1407)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G.,
            Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
            Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
            Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
            Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
            Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L.,
            Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koesema,E.,
            Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J.,
            Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1407
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="RAFL05-09-N10 (R10213)"
                     /note="This clone is in a modified pBluescript vector
                     (Lambda ZAP) as a XhoI/SstI insert.
                     ecotype: Columbia"
     gene            1. .1407
                     /gene="F24I3.90/AT3g57010"
     5'UTR           1. .97
                     /gene="F24I3.90/AT3g57010"
     CDS             98. .1228
                     /gene="F24I3.90/AT3g57010"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAK43996.1"
                     /db_xref="GI:13877837"
                     /translation="MPISRRVLTPITAAPVILAVLCFFFWSSIIGPDNLKGTKHVLQD
                     AKTIPLPVDGPESLEFDPQGEGPYVGVTDGRILKWRGEELGWVDFAYTSPHRDNCSSH
                     EVVPSCGRPLGLSFERKTGDLYICDGYFGVMKVGPEGGLGELVVDEAEGRKVMFANQG
                     DIDEEEDIFYFNDSSDTYHFRDVFYVSLSGTKVGRVIRYDMKKKEAKVIMDKLRLPNG
                     LALSKNGSFVVTCESSTNICHRIWVKGPKSGTNEVFATLPGSPDNIRRTPTGDFWVAL
                     HCKKNLFTRAVLIHTWVGRFFMNTMKMETVIHFMNGGKPHGIVVKLSGETGEILEILE
                     DSEGKTVKYVSEAYETKDGKLWIGSVYWPAVWVLDTSVYDSI"
     misc_difference 522
                     /gene="F24I3.90/AT3g57010"
                     /note="compared to genomic sequence resulting in an amino
                     acid sequence difference"
                     /replace="c"
     3'UTR           1229. .1407
                     /gene="F24I3.90/AT3g57010"
BASE COUNT      411 a    260 c    343 g    393 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AF370181/rev   ..

 Align seg 1/1  to reverse of: AF370181  from: 1  to: 1407

     144 ProAlaAsnThrGlnIleArgSer 151
         |||||||||||||||||||||||
    1187 CCGGCCAATACACAGATCCGATCC 1164

seq_name: gb_ov:IPJ00267

seq_documentation_block:
LOCUS       IPJ00267               1872 bp    mRNA    linear   VRT 28-DEC-1997
DEFINITION  Ictalurus punctatus mRNA for Oct1 transcription factor.
ACCESSION   AJ000267
VERSION     AJ000267.1  GI:2739211
KEYWORDS    transcription factor.
SOURCE      channel catfish.
  ORGANISM  Ictalurus punctatus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Siluriformes; Ictaluridae; Ictalurus.
REFERENCE   1  (bases 1 to 1872)
  AUTHORS   Ross,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1997) Ross D.A., Dept. of Biochemistry and
            Molecular Biology, Medical University of South Carolina, 171 Ashley
            Ave. Charleston, SC 29425-2211, USA
REFERENCE   2  (bases 1 to 1872)
  AUTHORS   Ross,D.A., Magor,B.G., Middleton,D.L., Wilson,M.R., Miller,N.W.,
            Clem,L.W. and Warr,G.W.
  TITLE     Oct1 transcription factor from the channel catfish
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .1872
                     /organism="Ictalurus punctatus"
                     /db_xref="taxon:7998"
                     /clone="97-54"
                     /cell_line="1B10"
                     /cell_type="B cell"
                     /clone_lib="lambda ZAPII"
     gene            1. .1872
                     /gene="Oct1"
     CDS             1. .1872
                     /gene="Oct1"
                     /codon_start=1
                     /product="Oct1 transcription factor"
                     /protein_id="CAA03984.1"
                     /db_xref="GI:2739212"
                     /db_xref="SPTREMBL:O57602"
                     /translation="MENGIHGVQAAQTNGLDSQQKPNGQATSAISNIQAQALLQQLSL
                     SSAQQQQLFLQQAQQILAAAVQQQQQQTANQQSSTTGANISASAATPITQIPLSQPIQ
                     IAHQLQPTGLNLQQFVLVQPGQLQPAQFIISPSPQGQTGLLQGLQAQNLLTQLPQSQA
                     HILQPQSSIALTTQAATPTRKIAAMPSAAPTTTKRVDVPLVEEASELEELEQFAKAFK
                     QRRIKLGFTQGDVGLAMGKLYGNDFSQTTISRFEALNLSFKNMCKLKPLLEKWLNDAE
                     NMTSESAVSSPSDMSPGICEGLNRRRKKRTSIDTNIRIALEKSFLEQSQKPTSEEITM
                     IADQLNMEKEVVRVWFCNRRQKEKRINPPSSSTSNTPIKTVFTNSSPSVISSSSNTLN
                     ASSALSLSTASVPNFTLAGTTLGPVTSNTATVISLAPAVATPAPTMTSPSLTPSATVA
                     TLTATKQEVVRTTETGSTIATSAAQGPGQFVMSPAGLGGANLAAMAAAAGLNPGIISP
                     SQLTQGGALLSLAPASLGNALTPTALMNNTTLATIQALASGGALPITSFDGNGNVLFA
                     NTSAGGPILFNSPLSLLPTGVLSAGGAGGGALNLQVTGDAHQSAMPVKSITVASKAQ"
BASE COUNT      487 a    560 c    447 g    378 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x IPJ00267   ..

 Align seg 1/1  to: IPJ00267  from: 1  to: 1872

     75 AlaThrLysGlnGluValValArg 82
        |||||||||||||||||||||||
   1360 GCTACAAAACAGGAAGTGGTCAGG 1383

seq_name: gb_pat:AR156086

seq_documentation_block:
LOCUS       AR156086               2007 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 7 from patent US 6241989.
ACCESSION   AR156086
VERSION     AR156086.1  GI:15124790
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2007)
  AUTHORS   Scott,F.W., Ngichabe,C.K., Hu,L. and Esposito,J.J.
  TITLE     Recombinant multivalent viral vaccine
  JOURNAL   Patent: US 6241989-A 7 05-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .2007
                     /organism="unknown"
BASE COUNT      528 a    473 c    421 g    585 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AR156086/rev   ..

 Align seg 1/1  to reverse of: AR156086  from: 1  to: 2007

    192 GluLeuGlnAsnArgLeuCysGln 199
        |||||||||||||||||||||||
   1090 GAATTACAAAATCGGTTATGTCAG 1067

seq_name: gb_vi:FCCPPKS19

seq_documentation_block:
LOCUS       FCCPPKS19              2010 bp    RNA     linear   VRL 18-NOV-1997
DEFINITION  Feline calicivirus gene encoding capsid protein precursor, strain
            KS109.
ACCESSION   X99446
VERSION     X99446.1  GI:1770254
KEYWORDS    capsid precursor protein.
SOURCE      Feline calicivirus.
  ORGANISM  Feline calicivirus
            Viruses; ssRNA positive-strand viruses, no DNA stage;
            Caliciviridae; Vesivirus.
REFERENCE   1  (bases 1 to 2010)
  AUTHORS   Geissler,K., Schneider,K., Platzer,G., Truyen,B., Kaaden,O.R. and
            Truyen,U.
  TITLE     Genetic and antigenic heterogeneity among feline calicivirus
            isolates from distinct disease manifestations
  JOURNAL   Virus Res. 48 (2), 193-206 (1997)
  MEDLINE   97318305
REFERENCE   2  (bases 1 to 2010)
  AUTHORS   Truyen,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1996) U. Truyen, Institute for Medical
            Microbiology, University of Muenich, Veteringer Str.13, Muenchen,
            D-80539, FRG
FEATURES             Location/Qualifiers
     source          1. .2010
                     /organism="Feline calicivirus"
                     /strain="KS109"
                     /db_xref="taxon:11978"
     CDS             1. .2010
                     /codon_start=1
                     /product="capsid protein precursor"
                     /protein_id="CAA67809.1"
                     /db_xref="GI:1770255"
                     /db_xref="SPTREMBL:P87673"
                     /translation="MCSTCANVLKYYEWDPHFKLVINPNIFLSVGFCDNPLMCCYPEL
                     LPEFGTVWDCNQSPLEIYLESILGDDEWSSTYEAIDPVVPPMHWGQAGKLFQPHPGVL
                     MHHLIKEVAKGWDPNLPSFRLEADGDSSVTSPEQGTMVGGVIAEPSAQMSTAADMATG
                     KSVDSEWEAFFSFHTSVNWSTSETQGKILFKQALGPLLNPYLEHLSQLYVAWSGSVEV
                     RFSISGSGVFGGKLAAIVVPPGVDPVQSTSMLQYPHVLFDARQVEPVIFTIPDLRSTL
                     YHLMSDTDTTSLVIMVYNDLINPYANESNSSGCIVTVETKPGSDFKFHLLKPPGSMLT
                     HGSVPSDLIPKSSSLWIGNRYWSDITDFVIRPFVFQANRHFDFNQETAGWSTPRFRPM
                     TVTMSVAKAAKLGIGIATDYIVPGIPDGWPDTTIPEDLVPAGDYAITNGKGNDITTAT
                     GYDTADIIRNNTNFRGMYICGSLQRAWGDKKISNTAFITTGIVEGNKLKPSNVIDQSK
                     IAIFQDTHANEEVQTSDDTLALLGYTGIGEEAIGSDRNKVVRISVLPEAGARGGNHPI
                     FYKNSIKLGYVIRSIDVFNSQILHTSRQLSLNHYLLSPDSFAVYRIIDANGSWFDIGI
                     DSDGFSFVGVSHIGKLEFPLSASYMGIQLAKIRLASNIRSTMTKL"
BASE COUNT      538 a    453 c    420 g    599 t
```

```
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x FCCPPKS19/rev   ..

 Align seg 1/1  to reverse of: FCCPPKS19  from: 1  to: 2010

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
    1093 GAATTACAAAATCGGTTATGTCAG 1070

seq_name: gb_pat:E12464

seq_documentation_block:
LOCUS       E12464                  2034 bp    DNA     linear   PAT 24-JUN-1998
DEFINITION  Capsid protein gene of feline calicivirus.
ACCESSION   E12464
VERSION     E12464.1  GI:3251297
KEYWORDS    JP 1997000267-A/2.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2034)
  AUTHORS   Kenjo,T., Maeda,T., Yokoyama,N. and Toya,Y. .
  TITLE     RECOMBINED FELINE HERPESVIRUS 1 TYPE CAPABLE OF MUTATING THYMIDINE
            KINASE GENE AND VACCINE CONTAINING THE SAME
  JOURNAL   Patent: JP 1997000267-A 2 07-JAN-1997;
            KYORITSU SHOJI KK
COMMENT     OS   Feline calicivirus
            PN   JP 1997000267-A/2
            PD   07-JAN-1997
            PF   23-JUN-1995 JP 1995179609
            PI   KENJO TAKESHI, MAEDA TAKESHI, YOKOYAMA NAOAKI, TOYA YUKINOBU
            PC   C12N15/09,A61K39/245,C12N5/10,C12N7/00,(C12N15/09,C12R1:92);
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .2034
            FT                   /organism='Feline calicivirus' FT   CDS
            19. .2025
            FT                   /product='capsid protein'.
FEATURES              Location/Qualifiers
     source           1. .2034
                      /organism="unidentified"
                      /db_xref="taxon:32644"
BASE COUNT      524 a    493 c    448 g    569 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x E12464/rev   ..

 Align seg 1/1  to reverse of: E12464  from: 1  to: 2034

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
    1108 GAATTACAAAATCGGTTATGTCAG 1085

seq_name: gb_pr:AK022355

seq_documentation_block:

LOCUS       AK022355                2077 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA FLJ12293 fis, clone MAMMA1001815.
ACCESSION   AK022355
VERSION     AK022355.1  GI:10433734
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens Mammary gland cDNA to mRNA, clone_lib:MAMMA1
            clone:MAMMA1001815.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
            Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
            Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
            Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2077)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES              Location/Qualifiers
     source           1. .2077
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="MAMMA1001815"
                      /clone_lib="MAMMA1"
                      /tissue_type="Mammary gland"
                      /note="cloning vector: pME18SFL3"
BASE COUNT      630 a    452 c    395 g    600 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AK022355/rev   ..

 Align seg 1/1  to reverse of: AK022355  from: 1  to: 2077

     211 GlyLeuGlyLeuTyrAlaThrGly 218
         |||||||||||||||||||||||
     937 GGACTAGGACTTTATGCCACTGGA 914

seq_name: gb_ba:AB010463

seq_documentation_block:
LOCUS       AB010463                2243 bp    DNA     linear   BCT 13-FEB-1999
DEFINITION  Vibrio parahaemolyticus gene for NorM, complete cds.
ACCESSION   AB010463
VERSION     AB010463.1  GI:3298335
KEYWORDS    NorM.
SOURCE      Vibrio parahaemolyticus (strain:AQ3334) DNA.
  ORGANISM  Vibrio parahaemolyticus
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 2243)
  AUTHORS   Tsuchiya,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JAN-1998) Tomofusa Tsuchiya, Okayama University,
```

```
            Faculty of Pharmaceutical Sciences; 1-1-1 Tsushima-naka, Okayama,
            Okayama 700-8530, Japan
            (E-mail:tsuchiya@pheasant.pharm.okayama-u.ac.jp,
            Tel:81-86-251-7957, Fax:81-86-251-7957)
  REFERENCE 2  (sites)
    AUTHORS Morita,Y., Kodama,K., Shiota,S., Mine,T., Kataoka,A., Mizushima,T.
            and Tsuchiya,T.
      TITLE NorM, a putative multidrug efflux protein, of Vibrio
            parahaemolyticus and its homolog in Escherichia coli
    JOURNAL Antimicrob. Agents Chemother. 42 (7), 1778-1782 (1998)
     MEDLINE 98325450
  FEATURES             Location/Qualifiers
       source          1. .2243
                       /organism="Vibrio parahaemolyticus"
                       /strain="AQ3334"
                       /db_xref="taxon:670"
       gene            833. .2203
                       /gene="norM"
       CDS             833. .2203
                       /gene="norM"
                       /codon_start=1
                       /transl_table=11
                       /product="NorM"
                       /protein_id="BAA31456.1"
                       /db_xref="GI:3298336"
                       /translation="MHRYKEEASSLIKLATPVLIASVAQTGMGFVDTVMAGGVTQTDM
                       AAVSVASSIWLPSILFGIGLLMALVPVVAQLNGSARREKIPFEIQQGVVLALLISIPI
                       IGVLLQTQFILQLMDVEAVMADKTVGYIHAVIFAVPAFLLFQTLRSFTDGMSLTKPAM
                       VIGFIGLLLNIPLNWIFVYGKFGAPELGGVGCGVATTIVYWVMFALLLAYVMTSSRLK
                       SINVFGEYHKPQWKAQVRLFKLGFPVAAALFFEVTLFAVVALLVSPLGPIIVAAHQVA
                       INFSSLVFMLPMSVGAAVSIRVGHRLGEENVDGARVASRVGIMVGLALATITAIITVL
                       SRELIAELYTNNPEVITLAMQLLLFAAVYQCTDAVQVIAAGALRGYKDMRAIFNRTFI
                       AYWILGLPTGYILGRTDWIVEPMGAQGFWLGFIIGLTAAALMLGVRLRWMHRQEPDVQ
                       LNFSLQ"
  BASE COUNT      527 a    537 c    503 g    676 t
  ORIGIN


  alignment_scores:
              Quality:    8.00            Length:      8
                Ratio:  1.000              Gaps:      0
   Percent Similarity: 100.000   Percent Identity: 100.000

  alignment_block:
   US-09-674-779-2 x AB010463   ..

   Align seg 1/1  to: AB010463  from: 1  to: 2243

       137 LeuLysSerArgGlyIleLeuPro 144
           ||||||||||||||||||||||||
       598 CTGAAGTCCCGCGGCATTTTGCCT 621

  seq_name: gb_vi:CVXCAPS

  seq_documentation_block:
  LOCUS       CVXCAPS                 2493 bp ss-RNA     linear   VRL 02-AUG-1993
  DEFINITION  Feline calicivirus capsid protein gene (put.), complete cds.
  ACCESSION   M32819
  VERSION     M32819.1  GI:323874
  KEYWORDS    capsid protein.
  SOURCE      Feline calicivirus (CFI/68 FIV), cDNA to viral RNA, clone pCV[2,7].
    ORGANISM  Feline calicivirus
              Viruses; ssRNA positive-strand viruses, no DNA stage;
              Caliciviridae; Vesivirus.
  REFERENCE   1  (bases 1 to 2493)
    AUTHORS   Neill,J.D., Reardon,I.M. and Heinrikson,R.L.
      TITLE   Nucleotide sequence and expression of the capsid protein gene of
              feline calicivirus
    JOURNAL   J. Virol. 65, 5440-5447 (1991)
     MEDLINE  91374597
  COMMENT     Draft entry and computer-readable sequence for [Unpublished (1990)]
              kindly submitted
              by J.D.Neill, 14-MAR-1990. Pioneer Hi-Bred International, 7300 NW
              62nd Ave., Johnston, IA 50131.
  FEATURES             Location/Qualifiers
       source          1. .2493
                       /organism="Feline calicivirus"
                       /db_xref="taxon:11978"
       CDS             127. .2133
                       /note="capsid protein ORF (put.); putative"
                       /codon_start=1
                       /protein_id="AAA42925.1"
                       /db_xref="GI:323875"
                       /translation="MCSTCANVLKYYDWDPHIKLVINPNKFLHVGFCDNPLMCCYPEL
                       LPEFGTMWDCDQSPLQVYLESILGDDEWSSTHEAIDPVVPPMHWDEAGKIFQPHPGVL
                       MHHLICKVAEGWDPNLPLFRLEADDGSITTPEQGTMVGGVIAEPNAQMSTAADMATGK
                       SVDSEWEAFFSFHTSVNWSTSETQGKILFKQSLGPLLNPYLTHLAKLYVAWSGSVDVR
                       FSISGSGVFGGKLAAIVVPPGIDPVQSTSMLQYPHVLFDARQVEPVIFSIPDLRSTLY
                       HLMSDTDTTSLVIMVYNDLINPYANDSNSSGCIVTVETKPGPDFKFHLLKPPGSMLTH
                       GSIPSDLIPKSSSLWIGNRFWSDITDFVIRPFVFQANRHFDFNQETAGWSTPRFRPIT
                       ITISVKESAKLGIGVATDYIVPGIPDGWPDTTIPGELVPVGDYAITNGTNNDITTAAQ
                       YDAATEIRNNTNFRGMYICGSLQRAWGDKKISNTAFITTGTVDGAKLIPSNTIDQTKI
                       AVFQDTHANKHVQTSDDTLALLGYTGIGEEAIGADRDRVVRISVLPERGARGGNHPIF
                       HKNSIKLGYVIRSIDVFNSQILHTSRQLSLNHYLLSPDSFAVYRIIDSNGSWFDIGID
                       NDGFSFVGVSSIGKLEFPLTASYMGIQLAKIRLASNIRSVMTKL"
       CDS             2130. .2450
                       /note="capsid protein ORF (put.); putative"
                       /codon_start=1
                       /protein_id="AAA42926.1"
                       /db_xref="GI:323876"
                       /translation="MNSILGLIDTVTNTIGKAQQIELDKAALGQQRELALQRMNLDRQ
                       ALNNQVEQFNKLLEQRVQGPIQSVRLARAAGFRVDPYSYTNQNFYDDQLNAIRLSYRN
                       LFKN"
  BASE COUNT      666 a    593 c    531 g    703 t
  ORIGIN


  alignment_scores:
              Quality:    8.00            Length:      8
                Ratio:  1.000              Gaps:      0
   Percent Similarity: 100.000   Percent Identity: 100.000

  alignment_block:
   US-09-674-779-2 x CVXCAPS/rev   ..

   Align seg 1/1  to reverse of: CVXCAPS  from: 1  to: 2493

       192 GluLeuGlnAsnArgLeuCysGln 199
           |||||||||||||||||||||||||
      1216 GAATTACAAAATCGGTTATGTCAG 1193

  seq_name: gb_ov:AF072657

  seq_documentation_block:
  LOCUS       AF072657                2927 bp    mRNA    linear   VRT 17-DEC-1998
  DEFINITION  Danio rerio HLH-containing transcription factor (Coe2) mRNA,
              complete cds.
  ACCESSION   AF072657
  VERSION     AF072657.1  GI:3273890
  KEYWORDS    .
  SOURCE      zebrafish.
    ORGANISM  Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
  REFERENCE   1  (bases 1 to 2927)
    AUTHORS   Bally-Cuif,L., Dubois,L. and Vincent,A.
      TITLE   Molecular cloning of zcoe2, the zebrafish homolog of Xenopus Xcoe2
              and mouse EBF-2, and its expression during primary neurogenesis
    JOURNAL   Mech. Dev. 77 (1), 85-90 (1998)
     MEDLINE  99002949
  REFERENCE   2  (bases 1 to 2927)
    AUTHORS   Bally-Cuif,L., Dubois,L. and Vincent,A.
      TITLE   Direct Submission
    JOURNAL   Submitted (16-JUN-1998) CNRS URA 1414, Ecole Normale Superieure, 46
              rue d'Ulm, Paris 75005, France
```

```
FEATURES             Location/Qualifiers
     source          1. .2927
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
     gene            1. .2927
                     /gene="Coe2"
     CDS             417. .2156
                     /gene="Coe2"
                     /note="similar to Xenopus laevis Xcoe2 and Mus musculus
                     EBF2"
                     /codon_start=1
                     /product="HLH-containing transcription factor"
                     /protein_id="AAC96103.1"
                     /db_xref="GI:3273891"
                     /translation="MFESQDHSIRTITELKVRTFEEEMDPVRSWVRNVGVIDANIAAQ
                     SGVALSRAHFEKQPPSNLRKSNFFHFVLALYDRNGQPVEVERTSFVDFVEQDKTGEKT
                     NNGTHYKLQLLYSNGVRTEQDLYARLIDSVTKQPISYEGQNKNPEMCRVLLTHEVMCS
                     RCCEKKSCGNRNETPSDPVIIDRFFLKFFLKCNQNCLKTAGNPRDMRRFQVVLSTTVC
                     VDGPVLAISDNMFVHNNSKHGRRSRRMDPNETVENNMEYATPCIKAISPSEGWTTGGA
                     MVIVIGENFFDGLQVVFGSMLVWSELITPHAIRVQTPPRHIPGVVEVTLSYKSKQFCK
                     GAPGRFIYTALNEPTIDYGFQRLQKLIPRHPGDPDKLAKEMLLKRAADVVESLYGNTT
                     SNQDMLLKRAADIAEALYSVPRPHSQLQAMPSSPVHGSVMGLSSYPTQLGVSIGEPGQ
                     TSGQGYTRNSSSLSPRGYPSSSTPQQSAYGSNGGMSYGAVPMSSLGVSGSPGFNSASP
                     NSSPYAIMPSSPPGSSSSSSLLPFSSFPSSTKQKSAFAPVLRPQGFPHHPSAKTSGGT
                     SFRAMTGLVVPPM"
BASE COUNT      811 a    646 c    658 g    812 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AF072657   ..

 Align seg 1/1  to: AF072657  from: 1  to: 2927

      98 MetSerGlnLeuLeuThrThrAla 105
         ||||||||||||||||||||||||
    1453 ATGAGCCAACTATTGACTACGGCT 1476

seq_name: gb_pr:AY032952

seq_documentation_block:
LOCUS       AY032952               3500 bp    mRNA    linear   PRI 08-MAY-2001
DEFINITION  Homo sapiens nGAP-like protein (AF9q34) mRNA, complete cds.
ACCESSION   AY032952
VERSION     AY032952.1  GI:14009345
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3500)
  AUTHORS   von Bergh,A.R.M., Wijers-Koster,P.M., Groot,A.J., Kluin,P.M. and
            Schuuring,E.
  TITLE     A novel Ras GTPase Activating Protein (RasGAP) gene is fused to MLL
            in acute myeloid leukemia with t(9;11)(q34;q23)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3500)
  AUTHORS   von Bergh,A.R.M., Wijers-Koster,P.M., Groot,A.J., Kluin,P.M. and
            Schuuring,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2001) Department of Pathology, Leiden University
            Medical Center, Albinusdreef 2, Leiden 2333 ZA, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .3500
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="9q34"
                     /note="acute myeloid leukemia patient"
     gene            1. .3500
                     /gene="AF9q34"
     CDS             148. .3345
                     /gene="AF9q34"
                     /note="MLL fusion partner"
                     /codon_start=1
                     /product="nGAP-like protein"
                     /protein_id="AAK50336.1"
                     /db_xref="GI:14009346"
                     /translation="MPRLKESRSHESLLSPSSAVEALDLSMEEEVVIKPVHSSILGQD
                     YCFEVTTSSGSKCFSCRSAAERDKWMENLRRAVHPNKDNSRRVEHILKLWVIEAKDLP
                     AKKKYLCELCLDDVLYARTTGKLKTDNVFWGEHFEFHNLPPLRTVTVHLYRETDKKKK
                     KERNSYLGLVSLPAASVAGRQFVEKWYPVVTPNPKGGKGPGPMIRIKARYQTITILPM
                     EMYKEFAEHITNHYLGLCAALEPILSAKTKEEMASALVHILQSTGKVKDFLTDLMMSE
                     VDRCGDNEHLIFRENTLATKAIEEYLKLVGQKYLQDALGEFIKALYESDENCEVDPSK
                     CSAADLPEHQGNLKMCCELAFCKIINSYCVFPRELKEVFASWRQECSSRGRPDISERL
                     ISASLFLRFLCPAIMSPSLFNLLQEYPDDRTARTLTLIAKVTQNLANFAKFGSKEEYM
                     SFMNQFLEHEWTNMQRFLLEISNPETLSNTAGFEGYIDLGRELSSLHSLLWEAVSQLE
                     QSIVSKLGPLPRILRDVHTALSTPGSGQLPGTNDLASTPGSGSSSISAGLQKMVIEND
                     LSGLIDFTRLPSPTPENKDLFFVTRSSGVQPSPARSSSYSEANEPDLQMANGGKSLSM
                     VDLQDARTLDGEAGSPAGPDVLPTDGQAAAAQLVAGWPARATPVNLAGLATVRRAGQT
                     PTTPGTSEGAPGRPQLLAPLSFQNPVYQMAAGLPLSPRGLGDSGSEGHSSLSSHSNSE
                     ELAAAAKLGSFSTAAEELARRPGELARRQMSLTEKGGQPTVPRQNSAGPQRRIDQPPP
                     PPPPPPPAPRGRTPPNLLSTLQYPRPSSGTLASASPDWVGPSTRLRQQSSSSKGDSPE
                     LKPRAVHKQGPSPVSPNALDRTAAWLLTMNAQLLEDEGLGPDPPHRDRLRSKDELSQA
                     EKDLAVLQDKLRISTKKLEEYETLFKCQEETTQKLVLEYQARLEEGEERLRRQQEDKD
                     IQMKGIISRLMSVEEELKKDHAEMQAAVDSKQKIIDAQEKRIASLDAANARLMSALTQ
                     LKERYSMQARNGISPTNPTKLQITENGEFRNSSNC"
BASE COUNT      769 a   1109 c   1057 g    565 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AY032952/rev   ..

 Align seg 1/1  to reverse of: AY032952  from: 1  to: 3500

     101 LeuLeuThrThrAlaArgSerTrp 108
         ||||||||||||||||||||||||
    2821 CTTCTAACAACTGCGCGTTCATGG 2798

seq_name: gb_vi:FCL3E

seq_documentation_block:
LOCUS       FCL3E                  3516 bp    RNA     linear   VRL 07-FEB-1999
DEFINITION  Feline calicivirus (FCV) genome, 3'-end, RNA polymerase and capsid
            precursor protein.
ACCESSION   D90357
VERSION     D90357.1  GI:221264
KEYWORDS    RNA polymerase; capsid precursor protein.
SOURCE      Feline calicivirus F4, cDNA to genomic RNA, clone pFCV119.
  ORGANISM  Feline calicivirus
            Viruses; ssRNA positive-strand viruses, no DNA stage;
            Caliciviridae; Vesivirus.
REFERENCE   1  (bases 1 to 3516)
  AUTHORS   Tohya,Y., Taniguchi,Y., Takahashi,E., Utagawa,E., Takeda,N.,
            Miyamura,K., Yamazaki,S. and Mikami,T.
  TITLE     Sequence analysis of the 3'-end of feline calicivirus genome
  JOURNAL   Virology 183 (2), 810-814 (1991)
  MEDLINE   91306470
COMMENT     Submittde (18-Mar-1991) to DDBJ by:                    Takeshi
            Mikami
            Department of Veterinary Microbiology
            Faculty of Agriculture, The University of Tokyo
            1-1-1 Yayoi, Bunkyo-ku
            Tokyo 113
            Japan
```

```
            Phone:  03-3812-2111 x5396
            Fax:    03-5689-7346.
FEATURES              Location/Qualifiers
     source           1. .3516
                      /organism="Feline calicivirus"
                      /db_xref="taxon:11978"
     CDS              <1. .1146
                      /note="RNA polymerase"
                      /codon_start=1
                      /protein_id="BAA14370.1"
                      /db_xref="GI:221265"
                      /translation="LDLLSSKWKLATQGIALPHEYTIGLKDELRPIEKMQEGKRRMIW
                      GCDVGVATVCAAAFKGVSDAITANHQYGPIQVGINMDSPSVEVLYQRIKSAAKVFAVD
                      FSNWDSTQSPRVSAASIDILRYFSDRSPIVDSAANTLKSPPIAIFNGVAVKVASGLPS
                      GMPLTSVINSLNHCMYVGCAILQSLEARQIPVTWNLFSSFDMMTYGDDGVYMFPTMFA
                      SVSDQIFGNLSAYGLKPTRVDKSVGAIESIDPESVVFLKRTITRTPNGIRGLLDRSSI
                      IRQFFYIKGENSDDWKTPPKTIDPTSRGQQLWNACLYASQHGVEFYNKVLKLAMRAVE
                      YEGLHLKPPSYSSALEHYNSQFNGVEARSDQINMSDVTALHCDVFEV"
     CDS              1149. .3155
                      /note="capsid precursor protein"
                      /codon_start=1
                      /protein_id="BAA14371.1"
                      /db_xref="GI:221266"
                      /translation="MCSTCANVLKYYDWDPHFRLIINPNKFLPIGFCDNPLMCCYPDL
                      LPEFGTVWDCDQSPLQIYLESILGDDEWASTHEAIDPSVPPMHWDSAGKIFQPHPGVL
                      MHHLIGEVAKAWDPNLPLFRLEADDGSITTPEQGTAVGGVIAEPSAQMSTAADMASGK
                      SVDSEWEAFFSFHTSVNWSTSETQGKILFKQSLGPLLNPYLEHLSKLYVAWSGSIEVR
                      FSISGSGVFGGKLAAIVVPPGVDPVQSTSMLQYPHVLFDARQVEPVIFTIPDLRSTLY
                      HVMSDTDTTSLVIMVYNDLINPYANDSNSSGCIVTVETKPGPDFKFHLLKPPGSVLTH
                      GSIPSDLIPKSSSLWIGNRYWTDITDFVIRPFVFQANRHFDFNQETAGWSTPRFRPIT
                      ITISEKNGSKLGIGVATDYIIPGIPDGWPDTTIADKLIPAGDYSITTGEGNDIKTAQA
                      YDTAAVVKNTTNFRGMYICGSLQRAWGDKKISNTAFITTAIRDGNEIKPSNTIDMTKL
                      AVYQDTHVEQEVQTSDDTLALLGYTGIGEEAIGSNRDRVVRISVLPEAGARGGNHPIF
                      YKNSIKLGYVIRSIDVFNSQILHTSRQLSLNHYLLPPDSFAVYRIIDSNGSWFDIGID
                      SEGFSFVGVSDIGKLEFPLSASYMGIQLAKIRLASNIRSRMTKL"
     CDS              3152. .3472
                      /note="short open reading frame"
                      /codon_start=1
                      /protein_id="BAA14372.1"
                      /db_xref="GI:221267"
                      /translation="MNSILGLIDTVTNTVGKGQQIELDKAALGQQRELALQRMSLDRQ
                      ALNNQVEQFNKLLEQRVHGPIQSVRLARAAGFRVDPYSYTNQNFYDDQLNAIRLSYRN
                      LFKN"
BASE COUNT      932 a    828 c    788 g    968 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x FCL3E/rev   ..

 Align seg 1/1  to reverse of: FCL3E  from: 1  to: 3516

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
    2238 GAATTACAAAATCGGTTATGTCAG 2215

seq_name: gb_pr:AB051530

seq_documentation_block:
LOCUS       AB051530               3790 bp    mRNA    linear   PRI 07-FEB-2001
DEFINITION  Homo sapiens mRNA for KIAA1743 protein, partial cds.
ACCESSION   AB051530
VERSION     AB051530.1  GI:12698030
KEYWORDS    .
SOURCE      Homo sapiens cDNA to mRNA, clone:pj01380.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Kikuno,R., Hattori,A., Kondo,Y., Okumura,K. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XIX. The complete sequences of 100 new cDNA clones from brain which
            code for large proteins in vitro
  JOURNAL   DNA Res. 7 (6), 347-355 (2000)
  MEDLINE   21082932
REFERENCE   2  (bases 1 to 3790)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Rsearch Institute,
            Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba -
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
            Fax:81-438-52-3914)
FEATURES              Location/Qualifiers
     source           1. .3790
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="pj01380"
                      /note="vector:pBluescriptII SK plus"
     gene             1. .1807
                      /gene="KIAA1743"
     CDS              <1. .1807
                      /gene="KIAA1743"
                      /note="Start codon is not identified."
                      /codon_start=2
                      /product="KIAA1743 protein"
                      /protein_id="BAB21834.1"
                      /db_xref="GI:12698031"
                      /translation="NLANFAKFGSKEEYMSFMNQFLEHEWTNMQRFLLEISNPETLSN
                      TAGFEGYIDLGRELSSLHSLLWEAVSQLEQSIVSKLGPLPRILRDVHTALSTPGSGQL
                      PGTNDLASTPGSGSSSISAGLQKMVIENDLSGLIDFTRLPSPTPENKDLFFVTRSSGV
                      QPSPARSSSYSEANEPDLQMANGGKSLSMVDLQDARTLDGEAGSPAGPDVLPTDGQAA
                      AAQLVAGWPARATPVNLAGLATVRRAGQTPTTPGTSEGAPGRPQLLAPLSFQNPVYQM
                      AAGLPLSPRGLGDSGSEGHSSLSSHSNSEELAAAAKLGSFSTAAEELARRPGELARRQ
                      MSLTEKGGQPTVPRQNSAGPQRRIDQPPPPPPPPPPAPRGRTPPNLLSTLQYPRPSSG
                      TLASASPDWVGPSTRLRQQSSSSKGDSPELKPRAVHKQGPSPVSPNALDRTAAWLLTM
                      NAQLLEDEGLGPDPPHRDRLRSKDELSQAEKDLAVLQDKLRISTKKLEEYETLFKCQE
                      ETTQKLVLEYQARLEEGEERLRRQQEDKDIQMKGIISRLMSVEEELKKDHAEMQAAVD
                      SKQKIIDAQEKRIASLDAANARLMSALTQLKESMH"
BASE COUNT      776 a   1181 c   1121 g    712 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AB051530/rev   ..

 Align seg 1/1  to reverse of: AB051530  from: 1  to: 3790

     101 LeuLeuThrThrAlaArgSerTrp 108
         |||||||||||||||||||||||||
    1370 CTTCTAACAACTGCGCGTTCATGG 1347

seq_name: gb_ro:AF236130

seq_documentation_block:
LOCUS       AF236130               6128 bp    mRNA    linear   ROD 22-AUG-2001
DEFINITION  Rattus norvegicus DOC2/DAB2 interactive protein mRNA, complete cds.
ACCESSION   AF236130
VERSION     AF236130.1  GI:15277524
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 6128)
```

```
  AUTHORS   Zhi,W., Tseng,C.-P. and Hsieh,J.-T.
  TITLE     A novel Ras GTPase activating protein that interacts with
            DOC-2/DAB2: a downstream effector leading to the suppression of
            prostate cancer
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6128)
  AUTHORS   Zhi,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2000) Urology, UT Southwestern Medical Center,
            5323 Harry Hines Blvd, Dallas, TX 75390-9110, USA
FEATURES             Location/Qualifiers
     source          1. .6128
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             458. .3448
                     /note="DIP1/2"
                     /codon_start=1
                     /product="DOC2/DAB2 interactive protein"
                     /protein_id="AAK93947.1"
                     /db_xref="GI:15277525"
                     /translation="MENLRRAVHPNKDNSRRVEHILKLWVIEAKDLPAKKKYLCELCL
                     DDVLYARTTGKLKTDNVFWGEHFEFHNLPPLRTVTVHLYRETDKKKKKERNSYLGLVS
                     LPAASVAGRQFVEKWYPVVTPNPKGGKGPGPMIRIKARYQTITILPMEMYKEFAEHIT
                     NHYLGLCAALEPILSAKTKEEMASALVHILQSTGKVKDFLTDLMMSEVDRCGDNEHLI
                     FRENTLATKGIEEYLKLVGHKYLQDALCEFIKALYESDENCEVDPSKCSAADLPEHQG
                     NLKMCCELAFCKIINSYCVFPRELKEVFASWRQECSSRGRPDISERLISASLFLRFLC
                     PAIMSPSLFNLLQEYPDDRTARTLTLIAKVTQNLANFAKFGSKEEYMSFMNQFLEHEW
                     TNMQRFLLEISNPETLSNTAGFEGYIDLGRELSSLHSLLWEAVSQLDQSIVSKLGPLP
                     RILRDVHTALSTPGSGQLPGTNDLASTPGSGSSSVSTGLQKMVIENDLSGLIDFTRLP
                     SPTPENKDLFFVTRSSGVQPSPARSSSYSEANEPDLQMANGSKSLSMVDLQDARTLDG
                     EAGSPVGPEALPADGQVPATQLVAGWPARAAPVSLAGLATVRRAVPTPTTPGTSEGAP
                     GRPQLLAPLSFQNPVYQMAAGLPLSPRGLGDSGSEGHSSLSSHSNSEELAAAAKLGSF
                     STAAEELARRPGELARRQMSLTEKGGQPTVPRQNSAGPQRRIDQPPPPPPPPPPAPRG
                     RTPPTMLSTLQYPRPSSGTLASASPDWAGPGTRLRQQSSSSKGDSPELKPRALHKQGP
                     SPVSPNALDRTAAWLLTMNAQLLEDEGLGPDPPHRDRLRSKEELSQAEKDLAVLQDKL
                     RISTKKLEEYETLFKCQEETTQKLVLEYQARLEEGEERLRRQQEDKDVQMKGIISRLM
                     SVEEELKKDHAEMQAAVDSKQKIIDAQEKRIASLDAANARLMSALTQLKERYSMRARN
                     GVSPTNPTKLQITENGEFRNSSNC"
BASE COUNT     1389 a    1791 c    1713 g    1235 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AF236130/rev   ..

 Align seg 1/1  to reverse of: AF236130  from: 1  to: 6128

     101 LeuLeuThrThrAlaArgSerTrp 108
         ||||||||||||||||||||||||
    2924 CTTCTAACAACTGCGCGTTCATGG 2901

seq_name: gb_pat:AX251205

seq_documentation_block:
LOCUS       AX251205               6291 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 173 from Patent WO0168912.
ACCESSION   AX251205
VERSION     AX251205.1  GI:15984628
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 6291)
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with tumor suppressor genes and
            oncogenes
  JOURNAL   Patent: WO 0168912-A 173 20-SEP-2001;
            Epigenomics AG (DE)
```

```
FEATURES             Location/Qualifiers
     source          1. .6291
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1501 a    210 c   1626 g   2954 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AX251205/rev   ..

 Align seg 1/1  to reverse of: AX251205  from: 1  to: 6291

      34 LysThrProSerValLeuIleThr 41
         |||||||||||||||||||||||
    1884 AAAACACCATCCGTTTTAATTACT 1861

seq_name: gb_vi:FCU13992

seq_documentation_block:
LOCUS       FCU13992               7677 bp    RNA     linear   VRL 17-APR-1998
DEFINITION  Feline calicivirus CFI/68 RNA helicase/cysteine
            protease/RNA-dependent RNA polymerase polyprotein precursor and
            capsid protein precursor, genes, complete cds; and unknown gene.
ACCESSION   U13992
VERSION     U13992.1  GI:3056875
KEYWORDS    .
SOURCE      Feline calicivirus.
  ORGANISM  Feline calicivirus
            Viruses; ssRNA positive-strand viruses, no DNA stage;
            Caliciviridae; Vesivirus.
REFERENCE   1  (bases 1 to 7677)
  AUTHORS   Neill,J.D.
  TITLE     Complete nucleotide sequence of feline calicivirus strain CFI/68
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7677)
  AUTHORS   Neill,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1994) VCR, National Animal Disease Center, P. O.
            Box 70, Ames, IA 50010, USA
REFERENCE   3  (bases 1 to 7677)
  AUTHORS   Neill,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1998) VCR, National Animal Disease Center, P. O.
            Box 70, Ames, IA 50010, USA
  REMARK    Sequence update by submitter
COMMENT     On Apr 17, 1998 this sequence version replaced gi:537254.
FEATURES             Location/Qualifiers
     source          1. .7677
                     /organism="Feline calicivirus"
                     /strain="CFI/68"
                     /specific_host="Felis domesticus"
                     /db_xref="taxon:11978"
     5'UTR           1. .19
     CDS             20. .5308
                     /codon_start=1
                     /product="RNA helicase/cysteine protease/RNA-dependent RNA
                     polymerase polyprotein precursor"
                     /protein_id="AAC13992.1"
                     /db_xref="GI:3056876"
                     /translation="MSQTLSFVLKTHNVRKDFVRSVKLTLARRRDLQYFYNKLSRSMR
                     AEACPSCASYDVCPNCTSSDIPDDGSSTELIPSWEEVTKTSTYSLLLSEDTSDELCPD
                     DLANVAAHIRKAISTQSHPANSDMCKEQLTSLLVMAEAMLPQRSRASIPLHQQHQAAR
                     LEWREKFFSKPLDFLLERIGVSKDILQITAIWKIILEKACYCKSYGEQWFTTAKQKLR
                     EMRSYESNTLKPLIGAFIDGLRFMTVDNPYPMGFLPKLIGLIKPLNLAMIIDNHENTL
                     SGWVITLTAIMELYNITECTIDLMTSLITAFYDKIGKATKFYSHVKALFTGFRTEDVS
                     NSFWYMAAAILCYLVTGLIPNNGRFLKIKCLLVRATTLVSGIIATQKLAAMFATWNSE
```

```
                     SIVNELSARTVAISELNNPTTTSDTESVERLLELAKILHEEIKVHTLNPIMQSYNLIL
                     RNLMSTLDGVITCCNKRKAIARKRQVPVCYILTGPPGCGKTTAAQALAKKLSDQEPSV
                     INLDVDHHDTYTGNEVCIIDEFDSSDKVDYANFVIGMVNSAPMVLNCDMLENKGKLFT
                     SKYIIMTSNSETPVKPSSKRAGAFYRRVTYHDVATLVESHKRARPGTAVPRSCYKKNF
                     SHLSLAKRGAECWCKEYVLDPKGLQHQSTKAPPPTFLNIDSLAQTMKQDFALKNMAFE
                     AEVGCSEHRYGFVCQQSEVETVRRLLNAIRMRLNATFTVCVGLEASNSVGCTAHVLTP
                     DEPFNGKRFVVSRCNEASLSALEGNCVQTALGVCMSNKDLTHLCHFIKGKIVNDSVRL
                     DELPANQHVVTVNSVFDLAWALRRHSTLTGQFQAIRAAYDVLHVPDKVPAMLRHWMDE
                     TSFSDEHVVTQFITPGGVVILESCGGARIWALGNNVIRAGGVTAIPTGGCVRLMGLSA
                     QTMPWSEILSELFSLLGKIWSSVKVSTLILTALSMYASRFRPKTEAKGKTKSKIGPYR
                     GRGVALTDDEYDEWKEHNAARKLDLSVEDFLMLRHRAALGADDTDAVKFRSWWNSRSR
                     LADDFEDVTVIGKGGVKHEKIRTNTLRAVDRGYDVSFAEESGPGAKFHKNAIGSVTDV
                     CGEHKGYCVHMGHGVYASVAHVVKGDSFFLGERIFDLKTNGEFCCFRSTKILPSAAPF
                     FSGRPTRDPWGSPVATDWKPKPYSTTSGKIVGCFATTSTETHPGDCGLPYIDDNGRVT
                     GLHTGSGGPKTPSAKLVVPYVHIDMKTKSVTAQKYDVTKPDISYKGLVCKQLDEIRII
                     PKGTRLHVSPAHLEDFEECSHQPASLGSGDPRCPKSLTAIVVDSLKPYCVVVNGPPHD
                     ILHRVQKMLIDHLSGFVPMNISSDTSMLSAFHKLNHDTSCGPYLGGRKKDHMVNGEPD
                     KALLDLLSSKWKLATQGIALPHEYTIGLKDELRPIEKVQEGKRRMIWGCDVGVATVCA
                     AAFKGVSDAITANHQYGPIQVGINMDSPSVEALFQRIKSARKVFAVDYSKWDSTQSPR
                     VSAASIDILRYFSDRTPIVDSATNTLKSPPIAVFNGVAVKVSSGLPSGMPLTSVINSL
                     NHCLYVGCAILQSLEARNVPVTWNLFSTFDMMTYGDDGVYMFPTMYASISDQIFANLS
                     AYGLKPTRVDKSVGSIEPIDPNSVVFLKRTITRTPQGIRGLLDRSSILRQFYYIKGEN
                     TDNWKEPPKTIDPMSRGQQLWNACLYASQHGIDFYNKVYKLAEKAVEYEGLHLEPPSY
                     STALEHYNSQFNGVEARTDQIDTSGMAALHCDVFEV"
     CDS             5311. .7317
                     /codon_start=1
                     /product="capsid protein precursor"
                     /protein_id="AAC13993.1"
                     /db_xref="GI:537256"
                     /translation="MCSTCANVLKYYDWDPHIKLVINPNKFLHVGFCDNPLMCCYPEL
                     LPEFGTMWDCDQSPLQVYLESILGDDEWSSTHEAIDPVVPPMHWDEAGKIFQPHPGVL
                     MHHLICKVAEGWDPNLPLFRLEADDGSITTPEQGTMVGGVIAEPNAQMSTAADMATGK
                     SVDSEWEAFFSFHTSVNWSTSETQGKILFKQSLGPLLNPYLTHLAKLYVAWSGSVDVR
                     FSISGSGVFGGKLAAIVVPPGIDPVQSTSMLQYPHVLFDARQVEPVIFSIPDLRSTLY
                     HLMSDTDTTSLVIMVYNDLINPYANDSNSSGCIVTVETKPGPDFKFHLLKPPGSMLTH
                     GSIPSDLIPKSSSLWIGNRFWSDITDFVIRPFVFQANRHFDFNQETAGWSTPRFRPIT
                     ITISVKESAKLGIGVATDYIVPGIPDGWPDTTIPGELVPVGDYAITNGTNNDITTAAQ
                     YDAATEIRNNTNFRGMYICGSLQRAWGDKKISNTAFITTGTVDGAKLIPSNTIDQTKI
                     AVFQDTHANKHVQTSDDTLALLGYTGIGEEAIGADRDRVVRISVLPERGARGGNHPIF
                     HKNSIKLGYVIRSIDVFNSQILHTSRQLSLNHYLLSPDSFAVYRIIDSNGSWFDIGID
                     NDGFSFVGVSSIGKLEFPLTASYMGIQLAKIRLASNIRSVMTKL"
     CDS             7314. .7634
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAC13994.1"
                     /db_xref="GI:537257"
                     /translation="MNSILGLIDTVTNTIGKAQQIELDKAALGQQRELALQRMNLDRQ
                     ALNNQVEQFNKLLEQRVQGPIQSVRLARAAGFRVDPYSYTNQNFYDDQLNAIRLSYRN
                     LFKN"
BASE COUNT     2141 a   1767 c   1677 g   2092 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x FCU13992/rev   ..

 Align seg 1/1  to reverse of: FCU13992  from: 1  to: 7677

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
    6400 GAATTACAAAATCGGTTATGTCAG 6377

seq_name: gb_vi:FCLF4

seq_documentation_block:
LOCUS       FCLF4                   7681 bp    RNA     linear   VRL 17-JUN-1999
DEFINITION  Feline calicivirus genomic RNA for non-structural proteins,
            complete cds.

ACCESSION   D31836
VERSION     D31836.1  GI:499895
KEYWORDS    non-structural protein.
SOURCE      Feline calicivirus (strain:F4) cDNA to genomic RNA, clone:clones
            pFCV[119, 148 and 208].
  ORGANISM  Feline calicivirus
            Viruses; ssRNA positive-strand viruses, no DNA stage;
            Caliciviridae; Vesivirus.
REFERENCE   1  (sites)
  AUTHORS   Oshikamo,R., Tohya,Y., Kawaguchi,Y., Tomonaga,K., Maeda,K.,
            Takeda,N., Utagawa,E., Kai,C. and Mikami,T.
  TITLE     The molecular cloning and sequence of an open reading frame
            encoding for non-structural proteins of feline calicivirus F4
            strain isolated in Japan
  JOURNAL   J. Vet. Med. Sci. 56 (6), 1093-1099 (1994)
  MEDLINE   95210407
REFERENCE   2  (bases 1 to 7681)
  AUTHORS   Tohya,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-1994) Yukinobu Tohya, Kagoshima University,
            Faculty of Agriculture, Department of Veterimary Microbiology;
            1-21-24 Koorimoto, Kagoshima, Kagoshima 890, Japan
            (Tel:0992-85-8725, Fax:0992-85-8725)
FEATURES             Location/Qualifiers
     source          1. .7681
                     /organism="Feline calicivirus"
                     /strain="F4"
                     /db_xref="taxon:11978"
                     /clone="clones pFCV[119, 148 and 208]"
     CDS             20. .5311
                     /note="ORF1 encoding the non-structural proteins"
                     /codon_start=1
                     /product="non-structural proteins"
                     /protein_id="BAA06622.1"
                     /db_xref="GI:1060879"
                     /translation="MSQTLSFVLKTHSVRKDFVHSVKVTLARRRDLQYLYNKLARTMR
                     AEACPSCSSYDVCPNCTSSDIPDNGSSTTSIPSWEDVTKTSTYSLLLSEDTSDELCPD
                     DLVNVAAHIRKALSTQAHPANTEMCKGQLTSSLVMAEAMLPQRSWASIPVHHQGLPAR
                     LEWREKLFFKPLGFLLGKIGVSKDTFQTTAIWKIILEKACYCKSYGDQWFCAAKQNLR
                     EMRTFESDTLKPLVGAFIDGLRFMTVDNPNPMGFLPKLIGLVKPLNLAMIIDNHENTL
                     SGWVVTLTAIMELYNITECTIDVITSLVTGFYDKISKATKFFSQVKALFTGFRSEDVG
                     NPFWYRAAANPCYLNTGLNPNNGRSPRIKACLSGATTLVSGIIATQKLAAMFATWNSE
                     SIVNELSARTVAISELNNPTTTSDTDSVERLLELAKILHEEIKVHTLNPIMQSYNPIL
                     RDLMSTFDGVITSCNKRKAIGRKRQVPVCYILTGPPGCGKTTAAQALAKKLSDQEPSV
                     INLDVDHHDTYTGNEVCIIDEFASPDKVDYANFVIGMVNFRPMVLNCDMLENKGKLFT
                     SKYIIMTSNSETPVKPSSKRAGAFYRRVTIIDVTNPLVESHKRARPGTSVPRSCYKKN
                     FSHLSLAKRGAECWCKEYVLDPKGLQHQTIKAPPPTFLNIDSLAQTMKQDFLLKNMAL
                     RTRDGCSEHRYGFICQQSEVETVRRLLNAIRARLNATFTVCVGPEASHSIGCTAHVLT
                     PDEPFNGRRFIVSRCNEASLAALEGNCVQSALGVCMSNKDLTHLCHFIRGKVVNDSVR
                     LDELPANKHVVTVNSVFDLAWAVRRHLTLTGQFQAIRAAYDVLTVPDKVPAMLRHWMD
                     ETSFSDEHVVTQFVTPGGVVILESCGGARIWALGHNVIRAGGVTATPTGGCVRLLGLS
                     AQTLPWSEIFRELFTLLDRIWSSIKVSTLVLTALGMYASRFRPKSEAKGKTKVKIGPY
                     RGRGVALTDDEYDEWREHNANRKLDLSVEDFLMLRHRAALGADDADAVKFRSWWNSRT
                     AQEMVLKTVTVIGKGGVKHEKIRTSTLRAVDRGYDVSFAEESGPGTKFHKNAIGSLTD
                     VCGEHKGYCVHMGHGVYASVAHVVKGDSYFLGERIFDVKTNGEFCCFRSTKILPSGAP
                     FFSGKPTRDPWGSPVATELKPKAYTTTSGKIVGCFATTSTETHPGDCGLPYIDDNGRV
                     TGLHTGSGGPKTPSAKLVVPYIHIDMKNKSVTPQKYDETKPNISYKGLVCKQLGEIRI
                     IPKGTRLHVSPAHVDGYQECSHQPLFRGQWRPRGSKSLTAIVVDSLKPYCDRVEGPPH
                     DVLHRVHEMLIDHLSGFVPMNIFLETFMYPAFHKLNHARSCGPYLGGRKKDHMVNGDP
                     DKKFLDLLSSKWKLATQGIALPHEYTIGLKDELRPIEKMQEGKRRMIWGCDVGVATVC
                     AAAFKGVSDAITANHQYGPIQVGINMDSPSVEVLYQRIKSAAKVFAVDFSNWDSTQSP
                     RVSAASIDILRYFSDRSPIVDSAANTLKSPPIAIFNGVAVKVASGLPSGMPLTSVINS
                     LNHCMYVGCAILQSLEARQIPVTWNLFSSFDMMTYGDDGVYMFPTMFASVSDQIFGNL
                     SAYGLKPTRVDKSVGAIESIDPESVVFLKRTITRTPNGIRGLLDRSSIIRQFFYIKGE
                     NSDDWKTPPKTIDPTSRGQQLWNACLYASQHGVEFYNKVLKLAMRAVEYEGLHLKPPS
                     YSSALEHYNSQFNGVEARSDQINMSDVTALHCDVFEV"
BASE COUNT     2046 a   1810 c   1777 g   2048 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x FCLF4/rev   ..

 Align seg 1/1  to reverse of: FCLF4  from: 1  to: 7681

     192 GluLeuGlnAsnArgLeuCysGln 199
         ||||||||||||||||||||||||
    6403 GAATTACAAAATCGGTTATGTCAG 6380

seq_name: gb_ba:AF400582

seq_documentation_block:
LOCUS       AF400582               8367 bp    DNA     linear   BCT 08-NOV-2001
DEFINITION  Acinetobacter sp. ADP1 mismatch repair protein (mutS), 7-Fe
            ferredoxin (fdxA), and O-methyltransferase-like protein genes,
            complete cds; and unknown genes.
ACCESSION   AF400582
VERSION     AF400582.1  GI:15217080
KEYWORDS    .
SOURCE      Acinetobacter sp. ADP1.
  ORGANISM  Acinetobacter sp. ADP1
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Acinetobacter.
REFERENCE   1  (bases 1 to 8367)
  AUTHORS   Young,D.M. and Ornston,L.N.
  TITLE     Functions of the Mismatch Repair Gene mutS from Acinetobacter sp.
            Strain ADP1
  JOURNAL   J. Bacteriol. 183 (23), 6822-6831 (2001)
  MEDLINE   21555098
   PUBMED   11698371
REFERENCE   2  (bases 1 to 8367)
  AUTHORS   Young,D.M. and Ornston,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2001) Molecular, Cellular, and Developmental
            Biology, Yale University, P.O. Box 208103, New Haven, CT
            06520-8103, USA
FEATURES             Location/Qualifiers
     source          1. .8367
                     /organism="Acinetobacter sp. ADP1"
                     /strain="ADP1; BD413"
                     /db_xref="taxon:62977"
     CDS             complement(12. .725)
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92492.1"
                     /db_xref="GI:15217081"
                     /translation="MTTLIQTPLSNAVNNFIDLNREKAEFAFEVFRETKTITANGTVN
                     FVERGPDNTLITFNYAGPWSKDKKSKVLVETLDGEVLLGRKTAGGRYNKLFREHVDVT
                     TISHVHAPNLGAWAQTHRTLPIRYVPVQRFHLFKELPVYIDRRQAEVDFILDQLKVDI
                     EYRAILEANGGATVWGRQGLQETAEFILLLEEGAQIQLLAELIGGSREYGQGVLHQQW
                     KMSGLTEKAKSLGLLPLNN"
     CDS             complement(739. .1797)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92493.1"
                     /db_xref="GI:15217082"
                     /translation="MSLVVNKKENSKEFVKNILSEIWVTRCPVPTATGISLKRGTLVK
                     KFEKHGIDVAVLQDARKGLSVHHFDHQLPALFREGGNVPALAARSEGAPSRLIGLTWI
                     DEWQVIIVRPDSDISQPEHLLNAKVALPEYADKRAGSIFRAMSLQGIRGALQLANLSF
                     RDVNFIEVPERAGRKENRDASQYWSGLDALITGQVDAVYVKGASAVEAATKRGLKVGI
                     DLDQFPDITSRVNNGTPRPITVHEDLLEDHFDVVVDFLEELFDTADWAQDHLDEVLTI
                     LQDETIAGAESVATAYRNNFHKSLHPSLSTERLELLKQQKDFLWLHGFLERDVNIDNW
                     VDHRPLLEVLKRRQNKQL"
     gene            2592. .5237
                     /gene="mutS"
     CDS             2592. .5237
                     /gene="mutS"
                     /note="MutS"
                     /codon_start=1
                     /transl_table=11
                     /product="mismatch repair protein"
                     /protein_id="AAK92494.1"
                     /db_xref="GI:15217083"
                     /translation="MNSTETMADLSSYTPMMQQYFKVKLEHQHALLFYRMGDFYELFF
                     DDARKAAKFLGITLTHRGKANGEPIPMAGVPYHAAEGYLARLVRAGQTVAICEQVGEG
                     ENAGSRCKAPMERKVVRIITPGTITDDALLGSYQSSNLVALCIQQNKIGIALLDLSAS
                     IFKVQQHEFKTEQLYIELARLMPSEIVVDEDLVDQNILEQIKKQIECSITKRPNVDFN
                     LNNAQKTLCDQFGVTTLSGFGIDHLPLAKAAAAALIHYAKETQKTALPHIRSIQLEQS
                     TDFIALDPITRRNLEIIDPLFEHGTSLFNLINDCQTAMGGRLLSRILMQPIRDTAILD
                     ARLDATEQLLIGYHESPVRLVLKEIGDIERVLSRVALGTARPRDLVQLRQACAQIPFL
                     RHALLPAIQTKSSKLLNELDHELGDFKDLHQLLLSAIVENPPVLLRDGNVIAEGYDAE
                     LDELRKIRDHAGQFLVDLEIKERESSGIPTLKIGYNRVSGYYIELTRAQAEQAPAYYI
                     RRQTLKNAERYITPELKSFEDKVLSSESRALAREKLLFESLLETLKKNIANLQMMSSA
                     IAQIDVLANFAHQARLQNWNRPDFSPEIGIKIHDGRHPVVEALSKTPYTPNDTFLDSQ
                     HRMAIITGPNMGGKSTFMRQTALISLLAYCGSYVPAKSAKLGPIDRIFTRIGSADDLS
                     TGKSTFMVEMTETSQILHHATHQSLVLMDEVGRGTSTYDGLSLAWACVLDLTKRIKCL
                     CLFATHYFELTELGSESAIDNYHVTAKEMNGNLILLHKVQHGPASQSHGLQVAKLAGI
                     PASVIKEAQKRLKILEKQQQQYFQTAVQNDLFAIPELPQIDETKIEIKKTSPVLDHLK
                     QLDVDSLTPRQALEALYQLKDQLEH"
     CDS             5493. .5942
                     /note="ORF1; similar to FAD-binding domain of AppA redox
                     regulator from Rhodobacter sphaerodies"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92495.1"
                     /db_xref="GI:15217084"
                     /translation="MAYVRLLYVSKTAKQHHEIKHDLMEILTTAIRFNSINKIKGVLY
                     YGNGYFVQCLEGEEKVIYDVFYNRIIKDSRHQNCEILSCKAVNELLFQKWSMKFAPLN
                     KKLRDFFQQYHIDEFNPYLLTHQTTEPFIELLSQEPHQEVDPYPHSS"
     gene            6267. .6590
                     /gene="fdxA"
     CDS             6267. .6590
                     /gene="fdxA"
                     /note="FdxA; similar to AvFdI from Azotobacter vinelandii"
                     /codon_start=1
                     /transl_table=11
                     /product="7-Fe ferredoxin"
                     /protein_id="AAK92496.1"
                     /db_xref="GI:15217085"
                     /translation="MTFVVTENCIKCKYQDCVEVCPVDCFYEGPNFLVINPDECIDCA
                     LCEPECPANAIFSEDELPEGQEVFIELNADLSQKWPNITQIGDQPADREEWNGKPDKL
                     QYLEK"
     CDS             6717. .7439
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92497.1"
                     /db_xref="GI:15217086"
                     /translation="MSQFLKGVLAISIIPIMMVGCTTAPQKKGKILDLPKGHRIHIPQ
                     ETVTIERPVRPKVTPNSYLNWLQNPSNRIQIQEYEQFLSRNDVANIVPNYELLKTARD
                     WQKCGRSEYAVPNRELWNNAIPTLRVLKYLVAAKILTDFEVTSVYRDLPLNQCAGGAN
                     SSRHLFNSAIDFRIGPEIPQPQDYAFIENTKFKLCQFWNQHGQSLNMGLGLYSSGQIH
                     IDTQGYRTWGPDHSRNSSMCNY"
     CDS             7593. .8261
                     /note="ORF3; member of O-methyltransferase family"
                     /codon_start=1
                     /transl_table=11
                     /product="O-methyltransferase-like protein"
                     /protein_id="AAK92498.1"
                     /db_xref="GI:15217087"
                     /translation="MQQLWTDIDHYIESNLIPEDSNLKFALQNTEDQGFPDHLAVAPN
                     QGMLLQMLIQMNQCKRVLELGTFAAYSTIWLARALPKDGYLMTIEGRDTHAALAQQNI
                     DHANLPVNIELKVGRAADILSKIDTDTFEPFDFIFIDADKQSYPEYLDLSLKLSHSGT
                     VIVLDNVIRAGALIDLDNQRPSIEGIRLTFEALQNHPQLLSCTALQTVGSKGHDGFAI
                     AIVK"
BASE COUNT     2656 a   1606 c   1445 g   2660 t
ORIGIN
```

```
alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AF400582   ..

 Align seg 1/1  to: AF400582  from: 1  to: 8367

     158 LeuAsnGlnCysAlaGlyGlyAla 165
         ||||||||||||||||||||||||
    7170 TTAAATCAGTGTGCTGGTGGCGCG 7193

seq_name: gb_ba:AE002287

seq_documentation_block:
LOCUS       AE002287               10780 bp    DNA     linear   BCT 26-MAY-2000
DEFINITION  Chlamydia muridarum, section 19 of 85 of the complete genome.
ACCESSION   AE002287 AE002160
VERSION     AE002287.2  GI:8163162
KEYWORDS    .
SOURCE      Chlamydia muridarum.
  ORGANISM  Chlamydia muridarum
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 10780)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
            pneumoniae AR39
  JOURNAL   Nucleic Acids Res. 28 (6), 1397-1406 (2000)
  MEDLINE   20150255
   PUBMED   10684935
REFERENCE   2  (bases 1 to 10780)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Jun 1, 2000 this sequence version replaced gi:7190237.
FEATURES             Location/Qualifiers
     source          1. .10780
                     /organism="Chlamydia muridarum"
                     /db_xref="taxon:83560"
                     /note="synonym: Chlamydia trachomatis MoPn"
     gene            complement(65. .364)
                     /gene="TC0200"
     CDS             complement(65. .364)
                     /gene="TC0200"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF73538.1"
                     /db_xref="GI:8163163"
                     /translation="MFKSQKPKKNKCCMWLRGVLFGGILATLLTSLFVPKSGIQLRKK
                     LLRVKTSGTKKGRALLRNSKHHTREFAEQTKTLAKNISKEIQDFTQSIIDESRRD"
     gene            complement(384. .746)
                     /gene="TC0201"
     CDS             complement(384. .746)
                     /gene="TC0201"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF39073.1"
                     /db_xref="GI:7190238"
                     /translation="MTPYLVETLVTCCILLSLGMVLIFAFAIGILLKLHKIMKHVSRL
                     ADLFNFEARWLAPLFLGKKFVLNWLRKKYNNRKISQLAEELTNDEDSEDSCSNRLIFG
                     VKLAAIALCAWLFIRKKD"
     gene            991. .2370
                     /gene="TC0202"
     CDS             991. .2370
                     /gene="TC0202"
                     /note="similar to GB:L12968 SP:P31120 GB:U01376 PID:290443
                     PID:606114; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoglucomutase/phosphomannomutase family
                     protein"
                     /protein_id="AAF39074.1"
                     /db_xref="GI:7190239"
                     /translation="MTCDVSPLFGTDGIRGRANFEPMTVETSVLLGKAVAGVLLERHS
                     GKHRVVVGKDTRLSGYMFENALIAGLTSMGIETLMLGPIPTPGVAFITRAYRADAGIM
                     ISASHNPYRDNGIKIFSSDGFKIGQAVEERIEAMIASKSFGGLPEDHAVGKNKRVKDA
                     MGRYIEYAKATFPKGRTLKGLRIVLDCAHGAAYRVAPSVFEELDAEVICYGCEPSGCN
                     INEGCGALWPSIIQKAVIEHEADVGIALDGDGDRLIMVDEKGHIVDGDMLLSICASDL
                     KRRQALPENRVVATVMTNFGVLRYLESLGIQVTISPVGDRHVLQHMLETKAILGGEQS
                     GHMIFLDYNTTGDGIVSALQVLRIMIESESTLSDLTACIVKSPQSLINVPVAKKVPLD
                     SLSNVQVVLQEVKEILGDSGRILLRYSGTENICRVMVEGTKKHQVDSLAKTIVDVVEA
                     EIGGAEISE"
     gene            2382. .4202
                     /gene="TC0203"
     CDS             2382. .4202
                     /gene="TC0203"
                     /note="similar to PID:1184044 SP:Q56213; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glucosamine--fructose-6-phosphate
                     aminotransferase (isomerizing)"
                     /protein_id="AAF39075.1"
                     /db_xref="GI:7190240"
                     /translation="MCGIFGYLGQRNAVPLVLEGLSKLEYRGYDSAGIAALTKDHLFV
                     EKSVGPVSQLCSKVSSDMHSQVAIGHTRWATHGEPSRFNAHPHVDMYESCALVHNGII
                     ENFQTLREDLSSKGVEFSSDTDTEVIVQLFASRYRETGDLVQSFSWTLKQLQGSFACA
                     LVHQDHPEILLCATYESPLILGLGEEEVFISSDVHAFLKYSCQIQTLASGELAVLRIG
                     RPVEIYNFELSRIQKEVRCIDHAEGSLDKQGFDYYMLKEIYEQPEVFERILHFVCEEN
                     GFAESFLKEFSFEGIESLHIVACGSSYHAGCLAKYVIESMVSIPVYVETASEFRYRQP
                     YIAKRSLAILISQSGETADTLAALNEFRKLDEVRVLGICNVRGSVLASRVDHCVFIEA
                     GLEVGVASTKAFTAQLLVLILLGLKLASQRQEISKQDLMQAVQGLRELPRLTRLFLDS
                     SIHDWRCRQSKETSFIFLGRRFMYPICMEAALKLKEIAYVEANAYPAGEMKHGPIALI
                     QEGTPVIVYCGDPFVYTKTIGAIMEVKARKAYVIALARESNQDIAAVSDEQIYIPDSH
                     DLAAPILFAIAGQIMAYTMALQKGTEVDRPRNLAKSVTVE"
     gene            4306. .5502
                     /gene="TC0204"
     CDS             4306. .5502
                     /gene="TC0204"
                     /note="similar to GB:M23240 SP:P18199 PID:148089 GB:U00096
                     PID:1736569; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="Mtr/TnaB/TyrO permease family protein"
                     /protein_id="AAF39076.1"
                     /db_xref="GI:7190241"
                     /translation="MINKMLGGALIVAGTTIGAGVLAVPIATSEGGFLPTTLLYVVSW
                     LIAVASGYCFLEVLTWLHARKNVNMVSMAECTLGYKSKVIMWLVYLLLFYSLLVAYFC
                     DGGNILMRVMGCRDWDTPWIRHAMPIVFFALFSPLLMAKTSIVDQCNRVFVLGLGISF
                     AMFCYFGFPLMKTELLVRSSWGGTLKGFPILFLAFGFQNVVPTLYHYMDKNVRDVKKA
                     IVIGSFIPLVLYVIWEAIVLGAVPVSFLEQAKMEGWTAIGALQGALKCSAFYIAGEFF
                     GFFALISSFIGVALGLKDFFIDAFQWDEKKRKLEIFLLVFVFPLVWAVFYPGIVLKCL
                     ECTGALGETIVLGVFPVLMVWKGRYGKKRYYGQRILPGGKGALLVMSGLVLVNLVLIV
                     QKFLGY"
     gene            5699. .6886
                     /gene="TC0205"
     CDS             5699. .6886
                     /gene="TC0205"
```

```
                     /note="similar to GB:M23240 SP:P18199 PID:148089 GB:U00096
                     PID:1736569; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="Mtr/TnaB/TyrO permease family protein"
                     /protein_id="AAF39077.1"
                     /db_xref="GI:7190242"
                     /translation="MRNKCIGGVLIVAGTVIGAGVLAVPVLTALDGFLPAALLYMLAW
                     LVSLASGYGYLEVLTWCKGNKQANLCSMAEETLGKVGRIVLCLVYLFLFYSLLVAYFC
                     DGGNILSRMLGDGVLENPWARHVMPILFFCIFAPLLMAKTSIIDYCNRVFVFGLILVF
                     GLFCILGAPRVQGDLLLRASWFSSLNSLPIFFLAFGFQNVVPSLYHYLDGDVREVKRA
                     IFIGSLIPLVLYVIWEALVLGTVPLVYLLKAKELGWTAAGALQGALKNSAFHIAGELF
                     GFFALVTSFIGTALALKDFYIDIFKWDARKQRLNLFLLVLVFPLVWAVSYPEIVLSCL
                     RYAGGIGGACIIVLFPVAMLWNGRYGKRHCSGKQILPGGKTVLLILLGYTVLNLAPLY
                     YTF"
gene                 6926. .7642
                     /gene="TC0206"
CDS                  6926. .7642
                     /gene="TC0206"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF39078.1"
                     /db_xref="GI:7190243"
                     /translation="MGLYDRDYTQDSRLPGTFSSRVYGWMTAGLAVTALTSLGLYATG
                     AYRTLFSLWWVWCFATLGVSFYIQAQIQKLSVPAVMGLFLAYSVLEGMFFGTMVPVYA
                     AQFGGGIVWAAFGSAAVIFGLSAAYGAFTKSDLTQIHRILMLALIGLMVISLGFLVVS
                     LFTPMPLMYLLICYLGLIIFVGLTVVDAQSIRRVARSVGDHGDLSYKLSLIMALQMYC
                     NVIMIFWYLLQIFASSDKRR"
gene                 complement(7705. .8559)
                     /gene="TC0207"
CDS                  complement(7705. .8559)
                     /gene="TC0207"
                     /note="similar to SP:P10121 GB:U00039 GB:X04398 PID:41498
                     PID:466600; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="signal recognition particle protein FtsY"
                     /protein_id="AAF39079.1"
                     /db_xref="GI:7190244"
                     /translation="MFKFFGNKLRSLFKKTLPSDLLEYAEALLYEGDFGPKLTEAFCN
                     ELRRHKHPDERLIKELIRSFLSDIIAKLPQKKSSSAHPHVTLILGTNGSGKTTTVAKL
                     AHYCLSQNQTVLIVATDTFRSAGMEQMRCWADSLGCGFISGKPGGDAAAIAFDGISAA
                     VARNYDQVIIDTSGRLHTHTNLLKELQKIATVCNKAFPGAPHETLMTVDATLGSNTLS
                     QVKLFHEAVPIDGLIFTKIEGSAKGGSLFRIADELKIPTRFVGYGETVYDFEPFSIDR
                     FLDKLLDV"
gene                 8629. .9789
                     /gene="TC0208"
CDS                  8629. .9789
                     /gene="TC0208"
                     /note="similar to GB:J01619 SP:P07460 PID:146203 GB:U00096
                     PID:1651323; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="succinyl-CoA synthetase, beta subunit"
                     /protein_id="AAF39080.1"
                     /db_xref="GI:7190245"
                     /translation="MHLHEYQAKDLLTSYQLPIPPYHVATSLSEAEVAIQAEQWRSGV
                     VKAQVHAGGRGKNGGVIVARSPEDLLAATDRLLHMQFSSNQTAGLSLPVNKVLISPLV
                     EIALEYYIAIVIDRKHRCPVIMLSKSGGIDIEEIAEKQPDLLLKIALPSSGKIYAYQL
                     RHIAKFMEWDKLVADRGNHIIRKLLQCFYDNDASLLEINPLVLTKDGDLIILDAKITI
                     DDNALYRHPQLADWYDPSQENIRDVLAKQLGLSYIALDGTIGCLVNGAGLAMSTLDIL
                     KLYGGSAANFLDVGGSASEKQIQEAISLVLSDKNVRVLFIHIFGGIMDCAVVASGLVS
                     AMQGQQGSIPTVIRLEGTNVDKGKEIILRSGIPCEFVASMSEGAELAVKLSR"
gene                 9804. .10679
                     /gene="TC0209"
CDS                  9804. .10679
                     /gene="TC0209"
                     /note="similar to GB:J01619 SP:P07459 PID:146204
                     PID:809679 GB:U00096; identified by sequence similarity;
```

```
alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AE002287   ..

 Align seg 1/1  to: AE002287  from: 1  to: 10780

     212 LeuGlyLeuTyrAlaThrGlyAla 219
         ||||||||||||||||||||||||
    7037 TTAGGGCTATATGCAACGGGAGCT 7060

seq_name: gb_ba:AE003962

seq_documentation_block:
LOCUS       AE003962              10986 bp    DNA     linear   BCT 15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 108 of 229 of the complete genome.
ACCESSION   AE003962 AE003849
VERSION     AE003962.1  GI:9106270
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10986)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 10986)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
```

```
          Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
          13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .10986
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(91. .1602)
                     /gene="XF1282"
     CDS             complement(91. .1602)
                     /gene="XF1282"
                     /note="similar to GI|6458686 (percent identity: 33 %/query
                     alignment coverage:  91.1 %/subject alignment coverage:
                     91.6 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="carboxypeptidase related protein"
                     /protein_id="AAF84091.1"
                     /db_xref="GI:9106271"
                     /translation="MAMIPRIALLSLLVSSSVPLAYATATPPKPDKETTTPATTLPAL
                     PANTQIRQSIRIGNRTLSYKAIIGALPVQDDKGKTIGEVMFTAYIMKGNNRPVTFALN
                     GGPGASSVYLNMGAIGPKVVHFGTQGTSPSDPAILHDNPGTWLDFTDLVFIDPIGTGF
                     SRARIEGEEAKKQFFNPKADVDYLSRTIYDWLVANGRLLSRKYLVGESYGGFRAPRIA
                     HTLQTQMGIALNGLVLVSPYLNPTLDDNSDVSPLAWIMTLPSIAAAHLERQHQLSEAK
                     MREVIEYTRGDYASTLMKGRSDPAATETMLQQVVHLTGLDPTYVRRAGGRLETQAYLR
                     EVFRDTGELGSRYDINVTALDPFPYDSEQRTNDPLLDSIIAPTTTAMVDFVTRTAGWK
                     VHERYQALNYDVNRQWDWNEELRKGSVTQLRQSVAIDPKLRVLIAHGWNDLSCPFMGS
                     VLTIDQIPKMGSDPTRIQVREYPGGHMFYNRPDSQSSFRQDVQAMYEHPQDTP"
     gene            1613. .1828
                     /gene="XF1283"
     CDS             1613. .1828
                     /gene="XF1283"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84092.1"
                     /db_xref="GI:9106272"
                     /translation="MVRNESSRGEQWGYQKLLEAMSSCWGWRAFGVGGLCVVRLYRDD
                     REVFVVDDLVAVCYRSASVTRMFRSID"
     gene            1821. .1979
                     /gene="XF1284"
     CDS             1821. .1979
                     /gene="XF1284"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84093.1"
                     /db_xref="GI:9106273"
                     /translation="MIDPNALWWILHGDCALPVSDVVVWFRGCIVQRASLLLSEPAVV
                     VLLALSME"
     gene            complement(2061. .2231)
                     /gene="XF1285"
     CDS             complement(2061. .2231)
                     /gene="XF1285"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84094.1"
                     /db_xref="GI:9106274"
                     /translation="MPITSPTRHAPKIFSPHNDIHQPSMHSMPHNQHTAQKTNVSFAM
                     HNYPLLLKTWLM"
     gene            complement(2338. .4227)
                     /gene="XF1286"
     CDS             complement(2338. .4227)
                     /gene="XF1286"
                     /note="similar to SP|P31598 (percent identity: 64 %/query
                     alignment coverage: 100.5 %/subject alignment coverage:
                     100.3 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="topoisomerase IV subunit B"
                     /protein_id="AAF84095.1"
                     /db_xref="GI:9106275"
                     /translation="MNSRYNAADIEVLTGLDPVKRRPGMYTDTTRPNHLAQEVIDNSV
                     DEALAGYANRIEVTLYKDGSCEVADNGRGMPVDIHPEEKIPGVELILTRLHAGGKFSN
                     RNYTFSGGLHGVGVSVVNALSSKVDLFIKREGSEHHMEFREGFAVSKLKVLGTVSKKN
                     TGTRLRFWPDSKYFDTPKFNVRVLRHLLRAKAVLCPGLTVELLEEATGEINTWHFKDG
                     LRDYLKGEMTESELLPAELFIGCLKKDTEVVDWAVAWVPEGELTQESYVNLIPTPQHG
                     THVNGLRSGLTDALREFCDFRNLLPRGIKLAPEDVWDRVTFVLSLKMSDPQFSGQTKE
                     RLSSRQAAGFIEGTVHDAFSLWLNQNVDIGEQIAQIAIERASARLKTEKQVTRKKITH
                     GPALPGKLADCISLDLSRTELFLVEGDSAGGSAKQARDKDFQAILPLRGKILNTWEVA
                     SGNVLASDEVHNLAIAIGCDPGKDNLSGLRYGKVVILADADSDGLHIATLLTALFLKH
                     FPALVAAGHVFVAMPPLFRVDVGKQVFYALDDEEKTMLLEKIKREKIKGQVSVTRFKG
                     LGEMNPQQLRESTIHPDTRRLVQLTMDDNTQTCSLMDMLLAKKRASDRKQWLETKGNL
                     ASLDI"
     gene            4261. .4701
                     /gene="XF1287"
     CDS             4261. .4701
                     /gene="XF1287"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84096.1"
                     /db_xref="GI:9106276"
                     /translation="MCCLLQIFARHSGSFSEGLAIPGRLGGGIPIEVALFRVRTGTSG
                     VATMKLKYLVMLMLVPAVVGVAVPSALAQAPKASSTATEMQQPASPSTDMPTNSEGQG
                     MGEHTGEHKARKSKKHRSHTVKPTSTGSATAPTKEAEENTSVNR"
     gene            4861. .6525
                     /gene="XF1288"
     CDS             4861. .6525
                     /gene="XF1288"
                     /note="similar to SP|P08398 (percent identity: 65 %/query
                     alignment coverage:  97.1 %/subject alignment coverage:
                     98.7 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="CTP synthetase"
                     /protein_id="AAF84097.1"
                     /db_xref="GI:9106277"
                     /translation="MTPLIFVTGGVVSSLGKGIAAASLASILEARGLKVTMVKLDPYI
                     NVDPGTMSPFQHGEVYVTDDGAETDLDLGHYERFVRTRLSRNNSVTTGRIYQNVICKE
                     RRGDYLGATVQVIPHITDEIRRCIDEATASFDVALVEIGGTVGDIESLPFLEAIRQVR
                     IERGAERTMFMHLTLVPYIAAAGELKTKPTQHSVKELRSIGIQPDVLLCRSEQVIPDS
                     ERRKIALFTNVSERAVIGCPDIDVLYGMPLELRRQGLDEIVIDQFKLSGTASLADLSE
                     WEDVVDAIKHPLDEVTIAVVGKYVDYQDAYKSVGEALKHGGLRQRTKVNLKWVEAQDL
                     EGSDMGALKDIDGILVPGGFGDRGFEGKVLASRYAREQRVPYFGICYGMQAAVVDYAR
                     HVAGLEGANSTENDRQSPHPVIALITEWRTTTGEVERRDEKSDLGGTMRLGLQEQRLK
                     AGTLVRELYGRDVVGERHRHRYEFNNRYRTQLEDAGLVIAAKSMDDTLVEMIELPREM
                     HPWFLACQAHPEFLSTPRDGHPLFIGFVKAARARKAGGKLLREVCV"
     gene            6629. .7516
                     /gene="XF1289"
     CDS             6629. .7516
                     /gene="XF1289"
                     /note="similar to SP|P17579 (percent identity: 44 %/query
                     alignment coverage:  94.6 %/subject alignment coverage:
                     98.2 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="2-dehydro-3-deoxyphosphooctonate aldolase"
                     /protein_id="AAF84098.1"
                     /db_xref="GI:9106278"
                     /translation="MRLSFFSVLLVIPKQSGDVMKLCGFEVGLNQPLFLIAGPCVIES

alignment_scores:
```

```
          Quality:    8.00            Length:      8
            Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AE003962/rev   ..

 Align seg 1/1  to reverse of: AE003962  from: 1  to: 10986

     143 LeuProAlaAsnThrGlnIleArg 150
         ||||||||||||||||||||||||
    1473 CTACCAGCCAACACACAGATACGG 1450

seq_name: gb_ba:AE001354

seq_documentation_block:
LOCUS       AE001354              11768 bp    DNA     linear   BCT 30-OCT-2000
DEFINITION  Chlamydia trachomatis section 81 of 87 of the complete genome.
ACCESSION   AE001354 AE001273
VERSION     AE001354.1  GI:3329280
KEYWORDS    .
SOURCE      Chlamydia trachomatis.
  ORGANISM  Chlamydia trachomatis
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 11768)
  AUTHORS   Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,
            Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
            Koonin,E.V. and Davis,R.W.
  TITLE     Genome sequence of an obligate intracellular pathogen of humans:
            Chlamydia trachomatis
  JOURNAL   Science 282 (5389), 754-759 (1998)
  MEDLINE   99000809
   PUBMED   9784136
REFERENCE   2  (bases 1 to 11768)
  AUTHORS   Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
            Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
  TITLE     Comparative genomes of Chlamydia pneumoniae and C. trachomatis
  JOURNAL   Nat. Genet. 21 (4), 385-389 (1999)
  MEDLINE   99206606
   PUBMED   10192388
REFERENCE   3  (bases 1 to 11768)
  AUTHORS   Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,
            Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
            Koonin,E.V. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1998) Program in Infectious Diseases, University
            of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES             Location/Qualifiers
     source          1. .11768
                     /organism="Chlamydia trachomatis"
                     /strain="D/UW-3/CX"
                     /db_xref="taxon:813"
     gene            153. .947
                     /gene="CT813"
     CDS             153. .947
                     /gene="CT813"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAC68409.1"
                     /db_xref="GI:3329281"
                     /translation="MTTLPNNCTSNSNSINTFTKDIEMAKQIQGSRKDPLAKTSWIAG
                     LICVVAGVLGLLAIGIGGCSMASGLGLIGAIVAAVIVAVGLCCLVSALCLQVEKSQWW
                     QKEFESWIEQKSQFRIVMADMLKANRKLQSEVEFLSKGWSDDTAVHKEDVTKYEQVVE
                     EYAEKIMELYEETGVLTIEKINLQKEKKAWLEEKAEMEQKLTTVTDLEAAKQQLEEKV
                     TDLESEKQELREELDKAIENLDEMAYEAMEFEKEKHGIKPGRRGSI"
     gene            complement(1031. .1432)
                     /gene="CT814"
     CDS             complement(1031. .1432)
                     /gene="CT814"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAC68410.1"
                     /db_xref="GI:3329282"
                     /translation="MQQPPTLWSQTCRYSPMRMALDKKKRLTLFVEKRMFRSQKPKKN
                     KCCLWLRGVLFGGFLATLLTSLFLPKSGMQIRKKLLRVKTSGTKKGRALLKNSKHHTR
                     EFAEQTKLLAKNISKEIQDFTQSIIDESRRD"
     gene            complement(1351. .1713)
                     /gene="CT814.1"
     CDS             complement(1351. .1713)
                     /gene="CT814.1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAC68411.1"
                     /db_xref="GI:3329283"
                     /translation="MTPYLIEALVACCILLSLGMVLIFAFTICLLLKLHKIMKLVSRI
                     SSLFNFEARWLAPLLIGKKFILNWLRKKYADRKMSQLADELEDNEDSENSCSSRLLCG
                     AKLAAIALCAWLLIRKKD"
     gene            1959. .3335
                     /gene="mrsA_2"
                     /note="CT815"
     CDS             1959. .3335
                     /gene="mrsA_2"
                     /codon_start=1
                     /transl_table=11
                     /product="Phosphoglucomutase"
                     /protein_id="AAC68412.1"
                     /db_xref="GI:3329284"
                     /translation="MTRDVRQLFGTDGVRGRANFEPMTVETSVLLGKAIAGVLLEKHA
                     GKHRVVVGKDTRLSGYMFENALIAGLTSMGIETLMLGPIPTPGVAFITRAYRADAGIM
                     ISASHNPYRDNGIKIFSSDGFKIGQAVEERIEAMVASKDFGKLPDDHAVGKNKRVKDA
                     TGRYIEYAKATFPKGRTLKGLRIVLDCAHGATYRVAPSVFEELDAEVICYGCEPSGCN
                     INAGCGALWPSTIQKAVIEHKADVGIALDGDGDRLIMVDEKGHIVDGDMLLSICASDL
                     KRRQALSDNRVVATVMTNFGVLRYLESLGIQVTISPVGDRHVLQHMLENQAVLGGEQS
                     GHMIFLDYNTTGDGIVSALQVLRIMIESESTLSDLTACIAKSPQALINVPVTKKVPLE
                     SLANVQGVLKEVKEVLGDSGRILLRYSGTENICRVMVEGTKKHQVDSLAKTIVDVVEA
                     EIGAEISE"
     gene            3346. .5166
                     /gene="glmS"
                     /note="CT816"
     CDS             3346. .5166
                     /gene="glmS"
                     /codon_start=1
                     /transl_table=11
                     /product="Glucosamine-Fructose-6-P Aminotransferase"
                     /protein_id="AAC68413.1"
                     /db_xref="GI:3329285"
                     /translation="MCGIFGYLGEKNAVPLVLEGLSKLEYRGYDSAGIATLVEGRLFV
                     EKAVGPVSQLCSAVSSDIHSQAAIGHTRWATHGEPSRFNAHPHVDMDASCALVHNGII
                     ENFQKLKEELEEQGVVFSSDTDTEVIVQLFARRYKETRDLIQSFSWTLKRLQGSFACA
                     LMHQDHPEVLLCAAHESPLILGLGEDEVFISSDIHAFLKYSGCTQTLASGELAVLRIG
                     KSIETYNFELARIQKEVRCIDHTEDSLDKKGFDYYMLKELYEQPEVFERILHLTCEEN
                     GFTESFLKGFSLDEIQSLHIVACGSSYHAGYLAKYVIESIASIPVYVETASEFRYRQP
                     YIAEHSLAILISQSGETADTLAALNEFRKLSKARVLGICNVRESALASRVDHCLFIEA
                     GLEVGVASTKAFTAQLLLLILLGLRLANHRQVIAQEDLAQAIQGLKDLPNLTRLFLDS
                     SIHDWRCRQIEETSFIFLGRRFMYPICMEAALKLKEIAYVEANAYPAGEMKHGPIALI
                     REGTPVIVYCGDRSVYTKTIGAIMEVKARKAYVIALAPESNRDIAAVSDEQIYIPDSH
                     DLAAPILFTIAGQIMAYTMALQRGTEVDRPRNLAKSVTVE"
     gene            5269. .6465
                     /gene="tyrP_1"
                     /note="CT817"
     CDS             5269. .6465
                     /gene="tyrP_1"
                     /codon_start=1
                     /transl_table=11
                     /product="Tyrosine Transport"
                     /protein_id="AAC68414.2"
                     /db_xref="GI:6578114"
                     /translation="MINKMLGGALLVAGTTIGAGVLAVPVSTSEGGFLPTTLLYIVSW
                     FIAVASGYCFLEVLTWTHSRKNVNMVSMAEYTLGHKSKIIMWLVYLLLFYSLLVAYFC
                     DGGNILMRVMGCRSWDTPWIRHAMPVVFFALFSPLLMAKTSIIDQCNRVFVFGLGIAF
                     AMFCYFGFPLMKTDLLVRSAWGATLKGFPILFLAFGFQNVVPTLYHYMDKNVKDVKKA
                     IVIGSSIPLVLYIIWEAIVLGAVPISFLEQAKVEGWTAIGALQTALKCSAFYVAGEFF
                     GFFALISSFIGVSLGLKDFFIDAFQWDEKKRKVEIFFLVFVFPLVWAVFYPGIVLKCL
```

```
                ECTGALGETIVLGVFPVLMVWKGRYGKKRYYGKRILPGGKGTLLVMSGLVLLNLVLIA
                QKFLGY"
     gene       6647. .7840
                /gene="tyrP_2"
                /note="CT818"
     CDS        6647. .7840
                /gene="tyrP_2"
                /codon_start=1
                /transl_table=11
                /product="Tyrosine Transport"
                /protein_id="AAC68415.1"
                /db_xref="GI:3329287"
                /translation="MCMRNKCVGGILIVAGTVIGAGVLAVPILTAVEGFFPAVVLYVL
                AWLVSLASGYGYLEVLTWCKGNRQANLCSMAEETLGRVGRIVLCLVYLFLFYSLLVAY
                FCDGGNILSRVIGESFFSYPWMRHVMPLLFFSLFAPLLMANTSVIDYCNRGFVFGLIF
                VFGLLCVLGVPRIQGELLLRASWFSSLNSLPIFFLAFGFQNVVPSLYHYLDGNIREVK
                RAILIGSLIPLILYIAWEALVLGTVPLVDLLKAKDLGWTAAGALQGSLKNSAFYIAGE
                LFGFFALVTSFIGTALALKDFYIDIFKWDARKKRVSLFFLVQVFPLVWAIFYPEIVLS
                CLRYAGGIGGACIIVLFPVAMLWNGRYGKRRCFGKRILPGGKTVLLILTGYTVLNLAT
                LYYTF"
     gene       7882. .8598
                /gene="yccA"
                /note="CT819"
     CDS        7882. .8598
                /gene="yccA"
                /codon_start=1
                /transl_table=11
                /product="Transport Permease"
                /protein_id="AAC68416.1"
                /db_xref="GI:3329288"
                /translation="MGLYDRDYAQDSRLPGTFSSRVYGWMTAGLAVTALTSLGLYATG
                AYRALFPMWWIWCFATLGVSFYIQAQIQKLSVPAVMGLFLAYSILEGMFFGTLVPVYA
                AQFGGGVVWAAFGSAGIIFGLSAAYGAFTKNDLTQIHRILMLALVGLVVISLAFLIVS
                LFTPMPLLYLLICYLGLIIFVGLTVVDAQSIRRVARSVGDHGDLSYKLSLIMALQMYC
                NVIMIFWYLLQIFASSDKRR"
     gene       complement(8696. .9550)
                /gene="ftsY"
                /note="CT820"
     CDS        complement(8696. .9550)
                /gene="ftsY"
                /codon_start=1
                /transl_table=11
                /product="Cell Division Protein FtsY"
                /protein_id="AAC68417.1"
                /db_xref="GI:3329289"
                /translation="MFKFFGNKLRSLFKKTLSSDLLEYAEVLLYEGDFGPKLTEAFCE
                ELRRCKNPDERAVKELIRSFLSKIISKLPQREPLSVRPFSTLVLGTNGSGKTTTVAKL
                AHYYLSQNQKVLIVATDTFRSAGMDQMRCWAETLNCGFISGKPGGDAAAIAFDGISAA
                VARDYDHVIIDTSGRLHTHTNLLKELQKIATVCNKAFPGAPHETLMTIDATLGSNTLS
                QVQLFHEAVPINGLIFTKVEGSAKGGSLFRIADELKIPTRFVGYGETIHDFEPFAIDR
                FLDKLLDV"
     gene       9620. .10780
                /gene="sucC"
                /note="CT821"
     CDS        9620. .10780
                /gene="sucC"
                /codon_start=1
                /transl_table=11
                /product="Succinyl-CoA Synthetase, Beta"
                /protein_id="AAC68418.1"
                /db_xref="GI:3329290"
                /translation="MHLHEYQAKDLLTAYQLPIPPYHVATSVPEVETAIQAEQWKAGV
                VKAQVHAGGRGKNGGVVIAHSPEDLLAAADKLLHMQFSSNQTAGLSLPINKVLISPLV
                EIASEYYLAIVIDRKHRCPVIMLSKAGGVDIEEVAEKQPDQLLKMTLPSSGKIYGYQL
                RRIAKFMEWDQPIADQGNRIIRQLLQCFYEKDASLLEINPLVLTKDGSLVILDAKMTI
                DDNALYRHPQLADCYDPSQENIRDVLAKQLGLSYIALDGTIGCLVNGAGLAMSTLDIL
                KLYGGSAANFLDVGGSASEKQIQEAISLVLSDKSVRVLFIHIFGGIMDCAVVASGLVS
                AMQGQKETIPTVIRLEGTNVDKGKDMIINAGIPCEFVTSMSEGAELAVQLSR"
     gene       10795. .11670
                /gene="sucD"

alignment_scores:
           Quality:    8.00          Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AE001354   ..

 Align seg 1/1  to: AE001354  from: 1  to: 11768

    212 LeuGlyLeuTyrAlaThrGlyAla 219
        ||||||||||||||||||||||||
   7993 CTAGGGTTATACGCAACAGGAGCG 8016

seq_name: gb_htg:AC110326

seq_documentation_block:
LOCUS       AC110326              18696 bp    DNA     linear   HTG 11-FEB-2002
DEFINITION  Rattus norvegicus clone CH230-296C10, *** SEQUENCING IN PROGRESS
            ***, 12 unordered pieces.
ACCESSION   AC110326
VERSION     AC110326.1  GI:18644766
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 18696)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 18696)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2002) Human Genome Sequencing Center, Department
```

```
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GQHV
              Center clone name: CH230-296C10
          -------------- Summary Statistics
              Sequencing vector: Plasmid; M77789
              Chemistry: Dye-terminator Big Dye: 98% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 18714 bases at least Q40
              Consensus quality: 20119 bases at least Q30
              Consensus quality: 20955 bases at least Q20
              Estimated insert size: 13754; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1805: contig of 1805 bp in length
          *     1806     1905: gap of unknown length
          *     1906     3339: contig of 1434 bp in length
          *     3340     3439: gap of unknown length
          *     3440     5444: contig of 2005 bp in length
          *     5445     5544: gap of unknown length
          *     5545     7473: contig of 1929 bp in length
          *     7474     7573: gap of unknown length
          *     7574     8607: contig of 1034 bp in length
          *     8608     8707: gap of unknown length
          *     8708    10633: contig of 1926 bp in length
          *    10634    10733: gap of unknown length
          *    10734    12022: contig of 1289 bp in length
          *    12023    12122: gap of unknown length
          *    12123    13353: contig of 1231 bp in length
          *    13354    13453: gap of unknown length
          *    13454    14523: contig of 1070 bp in length
          *    14524    14623: gap of unknown length
          *    14624    16078: contig of 1455 bp in length
          *    16079    16178: gap of unknown length
          *    16179    17401: contig of 1223 bp in length
          *    17402    17501: gap of unknown length
          *    17502    18696: contig of 1195 bp in length.
FEATURES             Location/Qualifiers
     source          1. .18696
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-296C10"
BASE COUNT     4861 a   3968 c   3724 g   5041 t   1102 others
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC110326/rev   ..

 Align seg 1/1  to reverse of: AC110326  from: 1  to: 18696

     101 LeuLeuThrThrAlaArgSerTrp 108
         ||||||||||||||||||||||||
    4151 CTACTAACAACAGCCAGAAGCTGG 4128

seq_name: gb_in:CEK07A1

seq_documentation_block:
LOCUS       CEK07A1                39086 bp    DNA     linear   INV 24-JAN-2002
DEFINITION  Caenorhabditis elegans cosmid K07A1, complete sequence.
ACCESSION   Z81097
VERSION     Z81097.1  GI:1729626
KEYWORDS    HTG; arginyl-tRNA protein transferase like; Chromatin assemble
            factor 1 P55 subunit like; Deoxyuridine 5'triphosphate
            nucleotidehydrolase; P58 protein like; Retinablastoma-binding
            protein RBAP46 like.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (sites)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
REFERENCE   2  (bases 1 to 39086)
  AUTHORS   Percy,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-1996) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Dec 12, 1996 this sequence version replaced gi:1628014.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            This sequence is the entire insert of clone K07A1. The true right
            end of clone F42H11 is at 16508 in this sequence. The start of this
            sequence (1. .105) overlaps with the end of sequence Z83225.
            The end of this sequence (38987. .39086) overlaps with the start of
            sequence AL023833.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=K07A1
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
FEATURES             Location/Qualifiers
     source          1. .39086
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="K07A1"
     gene            complement(join(1298. .1404,1453. .1585,1628. .1828,
                     1880. .2011,2204. .2289,2340. .2429,2480. .2609,2654. .2878))
                     /gene="K07A1.1"
     CDS             complement(join(1298. .1404,1453. .1585,1628. .1828,
                     1880. .2011,2204. .2289,2340. .2429,2480. .2609,2654. .2878))
                     /gene="K07A1.1"
                     /codon_start=1
                     /protein_id="CAB03174.1"
                     /db_xref="GI:3878338"
                     /db_xref="SPTREMBL:P90908"
                     /translation="MRKGKFLIVNESIEECVDRKLQVSTENCDVFPALVYIPKFTKKM
                     EKGMFCVLTIMAQILIYGVMGYFFKVGYLGLIVIEVAFLSVYGKEHVWLLRVQMIYLV
                     FQLVFSTLITFIAFLPDASGTRQDFSLIIPSLDRGTMALFGLTFGLFTCLRVILVASY
```

```
                HTHLFYFNLRDKFKDGKNISRIMMRGMFLVFNEPSDGNIEKKLDVSSQRCDVYPAHVY
                VPQMNKKAKSLYLLLCILDTLVHIASIGGRIFLKPDHYALYIFPGIMEVSFLVTFATK
                YIRFLHLQMIYLFAHTILSFIFLFANIVPNGKESLLVEFELPSLNSNTLIESVLAIFS
                FIRMLLVITYHQHVYYGELRQKFNQGKVIMQPQ"
gene            complement(join(3394. .3518,3591. .3680,3730. .3859,
                3908. .3994,4050. .4148))
                /gene="K07A1.13"
CDS             complement(join(3394. .3518,3591. .3680,3730. .3859,
                3908. .3994,4050. .4148))
                /gene="K07A1.13"
                /note="cDNA EST yk437d9.3 comes from this gene
                cDNA EST yk437d9.5 comes from this gene"
                /codon_start=1
                /protein_id="CAB03173.1"
                /db_xref="GI:3878337"
                /db_xref="SPTREMBL:Q9XVL4"
                /translation="MPSQTIFPKRTRRNRTIQSLNRLPEDDGDPTITGIAIILGIAEL
                IVSFLFLIFSQNYYFFAQLGLQLFFMALESYSLFVGFLVVAYAGAFEIIILLVFCGFE
                TFKPERSLIPTAFPGITETGRLWIIIAGYALVFWCGLRLIWTRRAMHYIYMARYEKRR
                NDALIAQMTAEGFVQV"
gene            join(4838. .5218,5265. .5386,5433. .5778,5828. .5949,
                5996. .6362,6414. .6551)
                /gene="K07A1.2"
CDS             join(4838. .5218,5265. .5386,5433. .5778,5828. .5949,
                5996. .6362,6414. .6551)
                /gene="K07A1.2"
                /note="Similarity to Human deoxyuridine 5'-triphosphate
                nucleotidohydrolase (SW:P33316), contains similarity to
                Pfam domain: PF00692 (dUTPase), Score=536.3,
                E-value=6.8e-158, N=3
                cDNA EST EMBL:M75928 comes from this gene; cDNA EST
                yk145c3.5 comes from this gene
                cDNA EST yk145c3.3 comes from this gene; cDNA EST
                yk636b3.3 comes from this gene
                cDNA EST yk636b3.5 comes from this gene; cDNA EST
                yk142g10.3 comes from this gene
                cDNA EST yk142g10.5 comes from this gene; cDNA EST
                yk504c3.3 comes from this gene
                cDNA EST yk480e5.3 comes from this gene; cDNA EST
                yk480e5.5 comes from this gene
                cDNA EST yk412g3.3 comes from this gene; cDNA EST
                yk412g3.5 comes from this gene
                cDNA EST yk471f10.3 comes from this gene; cDNA EST
                yk471f10.5 comes from this gene
                cDNA EST yk424e5.3 comes from this gene; cDNA EST
                yk424e5.5 comes from this gene
                cDNA EST yk413a9.3 comes from this gene; cDNA EST
                yk413a9.5 comes from this gene
                cDNA EST yk402h2.3 comes from this gene; cDNA EST
                yk402h2.5 comes from this gene
                cDNA EST yk399c6.3 comes from this gene; cDNA EST
                yk399c6.5 comes from this gene
                cDNA EST yk376e6.3 comes from this gene; cDNA EST
                yk376e6.5 comes from this gene
                cDNA EST yk373g11.3 comes from this gene; cDNA EST
                yk373g11.5 comes from this gene
                cDNA EST yk339d1.3 comes from this gene; cDNA EST
                yk339d1.5 comes from this gene
                cDNA EST yk282a2.3 comes from this gene; cDNA EST
                yk282a2.5 comes from this gene
                cDNA EST yk266d5.3 comes from this gene; cDNA EST
                yk266d5.5 comes from this gene
                cDNA EST yk281h10.5 comes from this gene; cDNA EST
                yk511b3.3 comes from this gene
                cDNA EST yk581g1.3 comes from this gene; cDNA EST
                yk596e12.3 comes from this gene
                cDNA EST yk611a5.3 comes from this gene; cDNA EST
                yk614a11.3 comes from this gene
                cDNA EST yk645b3.3 comes from this gene; cDNA EST
                yk667h7.3 comes from this gene
                cDNA EST yk674e1.3 comes from this gene; cDNA EST
                yk678h1.3 comes from this gene
                cDNA EST yk504c3.5 comes from this gene; cDNA EST
                yk511b3.5 comes from this gene
                cDNA EST yk581g1.5 comes from this gene; cDNA EST
                yk596e12.5 comes from this gene
                cDNA EST yk611a5.5 comes from this gene; cDNA EST
                yk614a11.5 comes from this gene
                cDNA EST yk645b3.5 comes from this gene; cDNA EST
                yk667h7.5 comes from this gene
                cDNA EST yk674e1.5 comes from this gene; cDNA EST
                yk678h1.5 comes from this gene"
                /codon_start=1
                /protein_id="CAB03175.1"
                /db_xref="GI:3878339"
                /db_xref="SPTREMBL:O01934"
                /translation="MTLTVAAPCVQTVTEPSKMEMQAESTSHITIRFTEMVGDAQKPT
                YGSISSAGADLYSAEDVVVPANGKLCVSTGLQIELPIGYYGRVAPRSGLAAKHFIDVG
                AGVIDSDYRGEVKVLLFNFNTTAFEVKTGDRIAKLICEQIGNGTYEEVKSLPSTNRGA
                GGFGSTGESTMNSETANPATNLERITVRFTQLNENAQTPTYGSEEAAGADLYSAEDIT
                VPAHGKCCVSTGIQMELPFGYYGRVAPRSGLAAKHFIDVGAGVIDSDYRGEVKVLLFN
                FTDNAFEVKKGDRIAQLICEKIGHCVYEAASELENTDRGAGGFGSTGQNTMEAEPALK
                KSKTEQIKTSATSVTIQITKSNDNAQMPTYGSAEAAGADLYSAEDVTVPARGKLCVST
                GIQMALPIGYYGRVAPRSGLAAKHFIDVGAGVIDSDYRGEVKVLLFNFGENDFEVKKG
                DRIAQLVCEQIALCTYSKVESLEVTERGAGGFGSTGQYITL"
gene            complement(join(6945. .7159,7203. .7335))
                /gene="K07A1.3"
CDS             complement(join(6945. .7159,7203. .7335))
                /gene="K07A1.3"
                /note="cDNA EST yk59e4.5 comes from this gene"
                /codon_start=1
                /protein_id="CAB03165.1"
                /db_xref="GI:3878329"
                /db_xref="SPTREMBL:P90910"
                /translation="MKVFRCGLVAISMCLEHFGIKTNPETILEKAKEMGFTKQGEMYS
                AESLASLTNEFLPNSSKIRKMPSPKEFTQSICDGKQMLIAYDCGPNYQPVYVRGHSAH
                WLLAVGTLIDSLN"
gene            complement(join(8137. .8319,8368. .8653,8809. .8966,
                9013. .9121,9180. .9331))
                /gene="K07A1.4"
CDS             complement(join(8137. .8319,8368. .8653,8809. .8966,
                9013. .9121,9180. .9331))
                /gene="K07A1.4"
                /codon_start=1
                /protein_id="CAB03166.1"
                /db_xref="GI:3878330"
                /db_xref="SPTREMBL:O01339"
                /translation="MHRSVIFLLTNVVLCIASVDRFLFDVSLTCHLHHIGRYSYKIQF
                VESDFWVLNKDDIITPPKFGEALPGSHNFTMFGKQEGDEITSEGYDVRMLLTHNCNGN
                REDIGAYMYVKPLCPIGTGLCHFKISRDITDMRGKVTITHKSFERMYGTTVFLLLTTV
                IIVVSSDDKFIFNVSFTCGYHQRESYSYNVQIFERDHWPLNGNDAITSQKIGKAHPGL
                SRFTMEGWQRSDEQFSKGYEPIIRLYHDCTEYKREAEVTLNITPSCEFGKGVCHYNIS
                KDIKDLAGEFPYKATLTYE"
gene            join(10040. .10132,10304. .10487,10537. .11060,11108. .11197)
                /gene="K07A1.5"
CDS             join(10040. .10132,10304. .10487,10537. .11060,11108. .11197)
                /gene="K07A1.5"
                /note="cDNA EST yk73g10.3 comes from this gene
                cDNA EST yk73g10.5 comes from this gene"
                /codon_start=1
                /protein_id="CAB03167.1"
                /db_xref="GI:3878331"
                /db_xref="SPTREMBL:O01340"
                /translation="MDVQDIIGIALNLATAGSLAAIILLPGCGGKKNRPAEGESSSSS

alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x CEK07A1/rev   ..

 Align seg 1/1  to reverse of: CEK07A1  from: 1  to: 39086

     153 TyrArgAsnProGluLeuAsnGln 160
```

```
          ||||||||||||||||||||||
   17484 TACCGTAATCCTGAATTAAATCAA 17461

seq_name: gb_pr:AL357372

seq_documentation_block:
LOCUS       AL357372               43553 bp    DNA     linear   PRI 20-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-86F19 on chromosome 10, complete
            sequence.
ACCESSION   AL357372
VERSION     AL357372.12  GI:15020748
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 43553)
  AUTHORS   Wilson,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 25, 2001 this sequence version replaced gi:14970346.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-86F19 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-86F19 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-399C16 is at 41554 in this
            sequence. The true right end of clone RP11-177H22 is at 2000 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .43553
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-86F19"
                     /clone_lib="RPCI-11.1"
     repeat_region   2. .213
                     /note="AluSg/x repeat: matches 144. .282 of consensus"
     repeat_region   213. .273
                     /note="MLT1D repeat: matches 128. .188 of consensus"
     repeat_region   814. .883
                     /note="35 copies 2 mer tt 65% conserved"
     repeat_region   1079. .1136
                     /note="29 copies 2 mer tt 72% conserved"
     repeat_region   1235. .1275
                     /note="MER4B repeat: matches 500. .539 of consensus"
     repeat_region   1385. .1623
                     /note="L2 repeat: matches 1312. .1543 of consensus"
     repeat_region   1698. .1944
                     /note="L2 repeat: matches 2040. .2289 of consensus"
     repeat_region   2015. .2036
                     /note="11 copies 2 mer aa 100% conserved"
     repeat_region   2100. .2400
                     /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region   2838. .3782
                     /note="L1M2 repeat: matches 5377. .6298 of consensus"
     repeat_region   3783. .4088
                     /note="AluSx repeat: matches 2. .307 of consensus"
     repeat_region   4089. .6514
                     /note="L1M2 repeat: matches 3010. .5377 of consensus"
     repeat_region   6515. .6801
                     /note="AluSx repeat: matches 3. .286 of consensus"
     repeat_region   6802. .6852
                     /note="L1M2 repeat: matches 2962. .3010 of consensus"
     repeat_region   6853. .7139
                     /note="AluSc repeat: matches 1. .287 of consensus"
     repeat_region   7140. .8465
                     /note="L1M2 repeat: matches 1669. .2962 of consensus"
     repeat_region   8466. .8758
                     /note="AluSg repeat: matches 1. .292 of consensus"
     repeat_region   8759. .8924
                     /note="L1M2 repeat: matches 1508. .1669 of consensus"
     repeat_region   8925. .9223
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   9224. .10543
                     /note="L1M2 repeat: matches -18. .1508 of consensus"
     repeat_region   10640. .11114
                     /note="L1M2 repeat: matches -677. .-191 of consensus"
     repeat_region   11200. .11249
                     /note="L1M2 repeat: matches 151. .200 of consensus"
     repeat_region   11278. .11373
                     /note="MIR repeat: matches 89. .187 of consensus"
     repeat_region   11793. .11947
                     /note="L2 repeat: matches 1404. .1573 of consensus"
     repeat_region   11988. .12104
                     /note="L2 repeat: matches 1825. .1931 of consensus"
     repeat_region   12105. .12405
                     /note="AluY repeat: matches 1. .302 of consensus"
     repeat_region   12406. .12774
                     /note="L2 repeat: matches 1931. .2351 of consensus"
     repeat_region   13807. .14032
                     /note="MIR repeat: matches 20. .257 of consensus"
     repeat_region   14146. .14445
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   14648. .14957
                     /note="AluJb repeat: matches 1. .312 of consensus"
     repeat_region   15479. .15615
                     /note="L2 repeat: matches 1663. .1809 of consensus"
     repeat_region   15650. .15767
                     /note="MIR repeat: matches 49. .178 of consensus"
     repeat_region   16056. .16346
                     /note="AluSx repeat: matches 1. .292 of consensus"
     repeat_region   16356. .16683
                     /note="AluJo repeat: matches 1. .310 of consensus"
     repeat_region   16707. .16958
                     /note="L2 repeat: matches 2466. .2710 of consensus"
     repeat_region   17120. .17340
                     /note="AluJb repeat: matches 87. .304 of consensus"
     repeat_region   17544. .17751
                     /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region   19449. .19746
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   21326. .21626
                     /note="AluSx repeat: matches 1. .296 of consensus"
     repeat_region   21890. .22192
                     /note="AluSp repeat: matches 1. .307 of consensus"
     repeat_region   22395. .22889
                     /note="11 copies 45 mer 71% conserved"
     repeat_region   22398. .22871
                     /note="158 copies 3 mer tct 71% conserved"
```

```
     repeat_region   22400. .22825
                     /note="213 copies 2 mer cc 54% conserved"
     misc_feature    22457. .22542
                     /note="Sequence from clone PCR only"
     repeat_region   22892. .23059
                     /note="AluSq repeat: matches 1. .168 of consensus"
     repeat_region   23625. .23934
                     /note="AluSg repeat: matches 3. .310 of consensus"
     repeat_region   24541. .24687
                     /note="MIR repeat: matches 47. .190 of consensus"
     repeat_region   25743. .25899
                     /note="MIR repeat: matches 79. .243 of consensus"
     repeat_region   26924. .26973
                     /note="25 copies 2 mer at 76% conserved"
     repeat_region   27039. .27156
                     /note="L1MD3 repeat: matches 7625. .7737 of consensus"
     repeat_region   30352. .30652
                     /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region   30668. .30965
                     /note="AluY repeat: matches 1. .297 of consensus"
     repeat_region   31412. .31543
                     /note="L1MA7 repeat: matches 6156. .6289 of consensus"
     repeat_region   31597. .31624
                     /note="14 copies 2 mer tt 96% conserved"
     repeat_region   33661. .33781
                     /note="FLAM_C repeat: matches 1. .121 of consensus"
     repeat_region   34496. .34693
                     /note="MIR repeat: matches 17. .216 of consensus"
     repeat_region   34897. .35244
                     /note="AluJb repeat: matches 1. .310 of consensus"
     repeat_region   36106. .36324
                     /note="MIR repeat: matches 3. .232 of consensus"
     repeat_region   38610. .38699
                     /note="2 copies 45 mer 96% conserved"
     repeat_region   40346. .40839
                     /note="L2 repeat: matches 2178. .2713 of consensus"
     repeat_region   40993. .41212
                     /note="L2 repeat: matches 1990. .2190 of consensus"
     repeat_region   41213. .41495
                     /note="AluSx repeat: matches 1. .287 of consensus"
     repeat_region   41496. .41983
                     /note="L2 repeat: matches 1392. .1990 of consensus"
     repeat_region   41996. .42088
                     /note="AluJb repeat: matches 136. .228 of consensus"
     repeat_region   42144. .42601
                     /note="L2 repeat: matches 732. .1323 of consensus"
     repeat_region   42903. .43194
                     /note="AluSg repeat: matches 1. .292 of consensus"
BASE COUNT    11508 a    8161 c    8761 g   15123 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AL357372/rev   ..

 Align seg 1/1  to reverse of: AL357372  from: 1  to: 43553

     168 SerLysHisLeuThrAsnSerAla 175
         ||||||||||||||||||||||||
   11182 AGTAAACATCTGACAAACAGTGCC 11159

seq_name: gb_htg:AC098224

seq_documentation_block:
LOCUS       AC098224               44428 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-34C16, *** SEQUENCING IN PROGRESS
            ***, 30 unordered pieces.
ACCESSION   AC098224
VERSION     AC098224.3  GI:17974394
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 44428)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 44428)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17062625.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIVE
                Center clone name: CH230-34C16
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 36210 bases at least Q40
                Consensus quality: 39629 bases at least Q30
                Consensus quality: 41587 bases at least Q20
                Estimated insert size: 11686; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation
```

```
              --------------
              * NOTE: Estimated insert size may differ from sequence length
              *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 30 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *        1     1856: contig of 1856 bp in length
              *     1857     1956: gap of unknown length
              *     1957     4072: contig of 2116 bp in length
              *     4073     4172: gap of unknown length
              *     4173     5702: contig of 1530 bp in length
              *     5703     5802: gap of unknown length
              *     5803     7463: contig of 1661 bp in length
              *     7464     7563: gap of unknown length
              *     7564     9576: contig of 2013 bp in length
              *     9577     9676: gap of unknown length
              *     9677    11737: contig of 2061 bp in length
              *    11738    11837: gap of unknown length
              *    11838    13410: contig of 1573 bp in length
              *    13411    13510: gap of unknown length
              *    13511    14639: contig of 1129 bp in length
              *    14640    14739: gap of unknown length
              *    14740    16113: contig of 1374 bp in length
              *    16114    16213: gap of unknown length
              *    16214    17474: contig of 1261 bp in length
              *    17475    17574: gap of unknown length
              *    17575    18813: contig of 1239 bp in length
              *    18814    18913: gap of unknown length
              *    18914    20077: contig of 1164 bp in length
              *    20078    20177: gap of unknown length
              *    20178    21581: contig of 1404 bp in length
              *    21582    21681: gap of unknown length
              *    21682    22975: contig of 1294 bp in length
              *    22976    23075: gap of unknown length
              *    23076    24104: contig of 1029 bp in length
              *    24105    24204: gap of unknown length
              *    24205    25456: contig of 1252 bp in length
              *    25457    25556: gap of unknown length
              *    25557    26776: contig of 1220 bp in length
              *    26777    26876: gap of unknown length
              *    26877    28191: contig of 1315 bp in length
              *    28192    28291: gap of unknown length
              *    28292    29372: contig of 1081 bp in length
              *    29373    29472: gap of unknown length
              *    29473    30582: contig of 1110 bp in length
              *    30583    30682: gap of unknown length
              *    30683    32269: contig of 1587 bp in length
              *    32270    32369: gap of unknown length
              *    32370    33466: contig of 1097 bp in length
              *    33467    33566: gap of unknown length
              *    33567    35103: contig of 1537 bp in length
              *    35104    35203: gap of unknown length
              *    35204    36237: contig of 1034 bp in length
              *    36238    36337: gap of unknown length
              *    36338    37837: contig of 1500 bp in length
              *    37838    37937: gap of unknown length
              *    37938    38961: contig of 1024 bp in length
              *    38962    39061: gap of unknown length
              *    39062    40129: contig of 1068 bp in length
              *    40130    40229: gap of unknown length
              *    40230    41866: contig of 1637 bp in length
              *    41867    41966: gap of unknown length
              *    41967    43065: contig of 1099 bp in length
              *    43066    43165: gap of unknown length
              *    43166    44428: contig of 1263 bp in length.
FEATURES             Location/Qualifiers
     source          1. .44428
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-34C16"
BASE COUNT     8998 a  11665 c  11518 g   9322 t   2925 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC098224/rev   ..

 Align seg 1/1  to reverse of: AC098224  from: 1  to: 44428

     138 LysSerArgGlyIleLeuProAla 145
         ||||||||||||||||||||||||
    3048 AAAAGTAGGGGTATCTTGCCAGCA 3025

seq_name: gb_htg:AC014969

seq_documentation_block:
LOCUS        AC014969               46275 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION   Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
             pieces.
ACCESSION    AC014969
VERSION      AC014969.1  GI:6436366
KEYWORDS     HTG; HTGS_PHASE2.
SOURCE       fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 46275)
  AUTHORS    Adams,M. and Venter,J.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
             Rockville, MD, USA
COMMENT      This sequence was identified as CDM:10210723 by the submitter.
             For further information on this sequence e-mail to fly@celera.com.
             * NOTE: This is a 'working draft' sequence.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .46275
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    12910 a  10354 c  10692 g  12319 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AC014969/rev   ..

 Align seg 1/1  to reverse of: AC014969  from: 1  to: 46275

       8 PheIleThrThrLeuIleSerSer 15
         ||||||||||||||||||||||||
   16423 TTCATTACTACGCTTATAAGCTCT 16400
```

```
OM of: US-09-674-779-2 to: Issued_Patents_NA:*   out_format : pfs

Date: Sep 16, 2002  7:56 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124144_26892/app_query.fasta_1.310
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=oli.rni -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09674779_@CGN1_1_47 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 45.910000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence          Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-19 -    8.00  152.83  0.7978    49  ! Paten
/cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-2 -     8.00  127.99  19.30     1242  ! Pate
/cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-3 -     8.00  127.86  19.61     1262  ! Pate
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-552-369-7 -    8.00  124.30  30.97     2007  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-743-637B-268 +    7.00  113.11  130.10    692  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-998-416-378 +    7.00  112.00  149.90    799  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-815-688A-4 -    7.00  109.35  210.57    1128  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-507-431-3 +    7.00  108.11  247.14    1327  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-700-546-1 +    7.00  108.11  247.14    1327  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-902-655A-3 +    7.00  108.11  247.14    1327  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-116-622-3 +    7.00  108.11  247.14    1327  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-219-277-3 +    7.00  108.11  247.14    1327  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-599-661-3 +    7.00  108.11  247.14    1327  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-000-630C-1 +    7.00  106.16  317.31    1710  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-862-730C-1 +    7.00  106.16  317.31    1710  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-202-056-4 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-076-093A-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-701-265-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-284-586-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-805-478-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-802-627A-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-238-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-228-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-104-296-3 +    7.00  106.04  322.25    1737  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-06380-2 +    7.00  106.04  322.25    1737
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-148-680-1 +    7.00  104.98  369.00    1993  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-522-6 -    7.00  104.33  400.92    2168  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-131-648-4 +    7.00  103.91  423.14    2290  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-176-620A-5 +    7.00  100.29  673.71    3671  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-463-862-5 +    7.00  100.29  673.71    3671  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-461-985-5 +    7.00  100.29  673.71    3671  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-458-887-5 +    7.00  100.29  673.71    3671  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5 +    7.00  100.29  673.71    3671  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-012C-5 +    7.00  100.29  673.71    3671  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-888-818C-5 +    7.00  100.29  673.71    3671  !
/cgn2_6/ptodata/2/ina/backfiles1.seq:5459251-1 +    7.00   98.19  882.01    4825  ! Pate
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-030-096-7 -    7.00   97.95  909.38    4977  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-485A-19 +    7.00   97.78  929.01    5086  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-365-486A-14 +    7.00   97.78  929.01    5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-080-285-19 +    7.00   97.78  929.01    5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-880-342-14 +    7.00   97.78  929.01    5086
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-05651-4 +    7.00   97.78  929.01    50
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-06251-2 +    7.00   97.78  929.01    50
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-234-186-7 +    7.00   97.77  930.45    5094
/cgn2_6/ptodata/2/ina/backfiles1.seq:5506344-1 +    7.00   97.75  932.25    5104  ! P

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-19

seq_documentation_block:
;Patent No. 5240845
;     APPLICANT: FUJII, SETSURO;TAKADA, KAORUKO;. KATANO, TAMIKI;
;MAJIMA, EIJI;OGINO, KOICHI;ONO, KENJI;SAKATA, YASUYO;UENOYAMA,
;TSUTOMU
;     TITLE OF INVENTION: MUTATED STREPTOKINASE PROTEINS
;     NUMBER OF SEQUENCES: 65
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/549,049
;       FILING DATE: 06-JUL-1990
;SEQ ID NO:19:
;       LENGTH: 49
5240845-19


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x 5240845-19/rev   ..

 Align seg 1/1  to reverse of: 5240845-19  from: 1  to: 49

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||
      26 GAGTTACAGAACCGTCTTTGTCAG 3

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-2

seq_documentation_block:
;Patent No. 5240845
;     APPLICANT: FUJII, SETSURO;TAKADA, KAORUKO;. KATANO, TAMIKI;
;MAJIMA, EIJI;OGINO, KOICHI;ONO, KENJI;SAKATA, YASUYO;UENOYAMA,
;TSUTOMU
;     TITLE OF INVENTION: MUTATED STREPTOKINASE PROTEINS
;     NUMBER OF SEQUENCES: 65
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/549,049
;       FILING DATE: 06-JUL-1990
;SEQ ID NO:2:
;       LENGTH: 1242
5240845-2


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x 5240845-2/rev   ..

 Align seg 1/1  to reverse of: 5240845-2  from: 1  to: 1242

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||
     379 GAGTTACAGAACCGTCTTTGTCAG 356

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-3

seq_documentation_block:
```

```
;Patent No. 5240845
;     APPLICANT: FUJII, SETSURO;TAKADA, KAORUKO;. KATANO, TAMIKI;
;MAJIMA, EIJI;OGINO, KOICHI;ONO, KENJI;SAKATA, YASUYO;UENOYAMA,
;TSUTOMU
;     TITLE OF INVENTION: MUTATED STREPTOKINASE PROTEINS
;     NUMBER OF SEQUENCES: 65
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/549,049
;       FILING DATE: 06-JUL-1990
;SEQ ID NO:3:
;       LENGTH: 1262
5240845-3


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x 5240845-3/rev    ..

 Align seg 1/1  to reverse of: 5240845-3  from: 1  to: 1262

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
     393 GAGTTACAGAACCGTCTTTGTCAG 370

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-552-369-7

seq_documentation_block:
; Sequence 7, Application US/08552369
; Patent No. 6241989
;  GENERAL INFORMATION:
;     APPLICANT:  Scott, Fred W.
;     APPLICANT:  Ngichabe, Christopher K.
;     APPLICANT:  Hu, Liangbiao
;     TITLE OF INVENTION:  Recombinant Multivalent Viral Vaccine
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hodgson, Russ, Andrews, Woods & Goodyear
;       STREET:  1800 One M&T Plaza
;       CITY:  Buffalo
;       STATE:  New York
;       COUNTRY:  United States
;       ZIP:  14203-2391
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 Kb storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS/ Microsoft Windows 3.1
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/552,369
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/190,789
;       FILING DATE:  01/27/1994
;       APPLICATION NUMBER:  07/726,609
;       FILING DATE:  07/09/1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Nelson, M. Bud
;       REGISTRATION NUMBER:  35,300
;       REFERENCE/DOCKET NUMBER:  18617.0016
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 856-4000
;       TELEFAX:  (716) 849-0349
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2007 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double-stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     HYPOTHETICAL:  yes
;     ORIGINAL SOURCE:
;       ORGANISM:  feline calicivirus
;     FEATURE:
;       LOCATION:  capsid protein gene region
US-08-552-369-7


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-552-369-7/rev    ..

 Align seg 1/1  to reverse of: US-08-552-369-7  from: 1  to: 2007

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
    1090 GAATTACAAAATCGGTTATGTCAG 1067

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-743-637B-268

seq_documentation_block:
; Sequence 268, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;     APPLICANT:  BERGERON, Michel G.
;     APPLICANT:  PICARD, Francois J.
;     APPLICANT:  OUELLETTE, Marc
;     APPLICANT:  ROY, Paul H.
;     TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;     TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;     TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;     TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;     NUMBER OF SEQUENCES:  273
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  QUARLES & BRADY
;       STREET:  411 EAST WISCONSIN AVENUE
;       CITY:  MILWAUKEE
;       STATE:  WISCONSIN
;       COUNTRY:  USA
;       ZIP:  53202-4497
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/743,637B
;       FILING DATE:  04-NOV-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/526,840
;       FILING DATE:  11-SEP-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BAKER, Jean C.
;       REGISTRATION NUMBER:  35,433
;       REFERENCE/DOCKET NUMBER:  850586.90012
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (414) 277-5000
;       TELEFAX:  (414)277-5591
;   INFORMATION FOR SEQ ID NO:  268:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  692 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Streptococcus mutans
US-08-743-637B-268
```

```
alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-743-637B-268    ..

 Align seg 1/1  to: US-08-743-637B-268  from: 1  to: 692

      58 SerHisValGlyLeuGlnAla 64
         ||||||||||||||||||||
     289 AGTCATGTTGGCTTACAAGCA 309

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-998-416-378

seq_documentation_block:
; Sequence 378, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  378:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  799 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1287RP
US-08-998-416-378


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-998-416-378    ..

 Align seg 1/1  to: US-08-998-416-378  from: 1  to: 799

      12 LeuIleSerSerMetLeuVal 18
         ||||||||||||||||||||
     490 TTAATATCTTCAATGCTTGTT 510

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-815-688A-4

seq_documentation_block:
; Sequence 4, Application US/08815688A
; Patent No. 5786195
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Shuang-yong
;    APPLICANT:  Xiao, Jian-ping
;    TITLE OF INVENTION:  METHOD FOR CLONING AND
;    TITLE OF INVENTION:  PRODUCING THE BSSHII RESTRICTION ENDONUCLEASE
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEW ENGLAND BIOLABS, INC.
;      STREET:  32 TOZER ROAD
;      CITY:  BEVERLY
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  01915
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/815,688A
;      FILING DATE:  12-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILLIAMS, GREGORY D
;      REGISTRATION NUMBER:  30901
;      REFERENCE/DOCKET NUMBER:  NEB-124
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-927-5054
;      TELEFAX:  508-927-1705
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1128 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...1125
;      OTHER INFORMATION:
US-08-815-688A-4


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-815-688A-4/rev    ..

 Align seg 1/1  to reverse of: US-08-815-688A-4  from: 1  to: 1128
```

```
     18 ValAlaCysSerAlaProIle 24
        ||||||||||||||||||||
    536 GTAGCTTGTTCAGCACCAATA 516

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-507-431-3

seq_documentation_block:
; Sequence 3, Application US/08507431
; Patent No. 5693518
;  GENERAL INFORMATION:
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Markus S.
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Heldt-Hansen, Hans P.
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Si, Joan Q.
;    APPLICANT:  Jacobson, Tina
;    APPLICANT:  Munk, Niels
;    APPLICANT:  Mullertz, Anette
;    TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACTIVITY FROM
;    TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5693518o No. 5693518disk of No. 5693518th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/507,431
;      FILING DATE:  15-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/002,800
;      FILING DATE:  25-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  3954.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4..1221, 1225..1314, 1318..1326)
US-08-507-431-3


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-507-431-3    ..

 Align seg 1/1  to: US-08-507-431-3  from: 1  to: 1327

      60 ValGlyLeuGlnAlaHisPhe 66
         ||||||||||||||||||||||
     676 GTAGGGTTGCAGGCCCATTTC 696

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-700-546-1

seq_documentation_block:
; Sequence 1, Application US/08700546
; Patent No. 5874274
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobsen, Tina Sejersgard
;    APPLICANT:  Heldt-Hansen, Hans Peter
;    APPLICANT:  Kofod, Lene Venke
;    APPLICANT:  Bagger, Christian
;    APPLICANT:  Mullertz, Anette
;    TITLE OF INVENTION:  Process For Processing Plant Material
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5874274o No. 5874274disk of No. 5874274th America, Inc.
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/700,546
;      FILING DATE:  23-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4141.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4..1221, 1225..1314, 1318..1326)
US-08-700-546-1


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-700-546-1    ..

 Align seg 1/1  to: US-08-700-546-1  from: 1  to: 1327

      60 ValGlyLeuGlnAlaHisPhe 66
         ||||||||||||||||||||||
     676 GTAGGGTTGCAGGCCCATTTC 696

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-902-655A-3

seq_documentation_block:
; Sequence 3, Application US/08902655A
; Patent No. 5885819
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Markus S.
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Heldt-Hansen, Hans P.
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Si, Joan Q.
;    APPLICANT:  Jacobson, Tina
;    APPLICANT:  Munk, Niels
;    APPLICANT:  Mullertz, Anette
;    TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACTIVITY FROM
;    TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5885819o No. 5885819disk of No. 5885819th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/902,655A
;      FILING DATE:  30-July-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris, Cheryl T.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  3954.214-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4..1221, 1225..1314, 1318..1326)
US-08-902-655A-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-902-655A-3    ..

 Align seg 1/1  to: US-08-902-655A-3  from: 1  to: 1327

      60 ValGlyLeuGlnAlaHisPhe 66
         ||||||||||||||||||||
     676 GTAGGGTTGCAGGCCCATTTC 696

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-116-622-3

seq_documentation_block:
; Sequence 3, Application US/09116622
; Patent No. 6080567
;  GENERAL INFORMATION:
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Markus S.
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Heldt-Hansen, Hans P.
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Si, Joan Q.
;    APPLICANT:  Jacobson, Tina
;    APPLICANT:  Munk, Niels
;    APPLICANT:  Mullertz, Anette
;    TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACTIVITY FROM
;    TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6080567o No. 6080567disk of No. 6080567th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/116,622
;      FILING DATE:  16-July-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris, Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  3954.224-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4..1221, 1225..1314, 1318..1326)
US-09-116-622-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-116-622-3    ..

 Align seg 1/1  to: US-09-116-622-3  from: 1  to: 1327

      60 ValGlyLeuGlnAlaHisPhe 66
         ||||||||||||||||||||
     676 GTAGGGTTGCAGGCCCATTTC 696

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-219-277-3

seq_documentation_block:
; Sequence 3, Application US/09219277
; Patent No. 6197564
;  GENERAL INFORMATION:
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Markus S.
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Heldt-Hansen, Hans P.
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Si, Joan Q.
;    APPLICANT:  Jacobson, Tina
```

```
;    APPLICANT:  Munk, Niels
;    APPLICANT:  Mullertz, Anette
;    TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACTIVITY FROM
;    TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6197564o No. 6197564disk of No. 6197564th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/219,277
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/116,622
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris, Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  3954.224-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4..1221, 1225..1314, 1318..1326)
US-09-219-277-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-219-277-3    ..

 Align seg 1/1  to: US-09-219-277-3  from: 1  to: 1327

      60 ValGlyLeuGlnAlaHisPhe 66
         ||||||||||||||||||||
     676 GTAGGGTTGCAGGCCCATTTC 696

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-599-661-3

seq_documentation_block:
; Sequence 3, Application US/09599661
; Patent No. 6228630
;  GENERAL INFORMATION:
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Markus S.
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Heldt-Hansen, Hans P.
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Si, Joan Q.
;    APPLICANT:  Jacobson, Tina
;    APPLICANT:  Munk, Niels
;    APPLICANT:  Mullertz, Anette
;    TITLE OF INVENTION:  ENZYMES WITH XYLANASE ACTIVITY FROM
;    TITLE OF INVENTION:  ASPERGILLUS ACULEATUS
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6228630o No. 6228630disk of No. 6228630th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/599,661
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/116,622
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Agris, Cheryl H.
;      REGISTRATION NUMBER:  34,086
;      REFERENCE/DOCKET NUMBER:  3954.224-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4..1221, 1225..1314, 1318..1326)
US-09-599-661-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-599-661-3    ..

 Align seg 1/1  to: US-09-599-661-3  from: 1  to: 1327

      60 ValGlyLeuGlnAlaHisPhe 66
         ||||||||||||||||||||
     676 GTAGGGTTGCAGGCCCATTTC 696

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-000-630C-1

seq_documentation_block:
; Sequence 1, Application US/09000630C
; Patent No. 6018029
;  GENERAL INFORMATION:
;    APPLICANT:  Fuller, Gerald M
;    APPLICANT:  Fuentes, Nelson L.
;    TITLE OF INVENTION:  DNA Encoding Canine Interleukin-1 Receptor
;    TITLE OF INVENTION:  Antagonist
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Douglas C Murdock/ Bradley, Arant, Rose & White
;      STREET:  2001 Park Place, Suite 1400
;      CITY:  Birmingham
;      STATE:  Alabama
```

```
;      COUNTRY:  USA
;      ZIP:  35203-2736
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Microsoft Windows
;      SOFTWARE:  WordPerfect 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/000,630C
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/862,730
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1710 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      CELL TYPE:  canine peripheral blood macrophage
;      CELL LINE:  primary monocytes
;    IMMEDIATE SOURCE:
;      LIBRARY:  lambda gt11 cDNA
;      CLONE:  Canine IL-1ra
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1 to 1710
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  open reading frame
;      LOCATION:  60 to 587
;      OTHER INFORMATION:
US-09-000-630C-1


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-000-630C-1    ..

 Align seg 1/1  to: US-09-000-630C-1  from: 1  to: 1710

      31 SerProIleLysThrProSer 37
         ||||||||||||||||||||
    1358 TCTCCCATCAAAACTCCCAGC 1378

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-862-730C-1

seq_documentation_block:
; Sequence 1, Application US/08862730C
; Patent No. 6063600
;  GENERAL INFORMATION:
;    APPLICANT:  Fuller, Gerald M
;    APPLICANT:  Fuentes, Nelson L.
;    TITLE OF INVENTION:  DNA Encoding Canine Interleukin-1 Receptor
;    TITLE OF INVENTION:  Antagonist
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Douglas C Murdock/ Bradley, Arant, Rose & White
;      STREET:  2001 Park Place, Suite 1400
;      CITY:  Birmingham
;      STATE:  Alabama
;      COUNTRY:  USA
;      ZIP:  35203-2736
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Microsoft Windows
;      SOFTWARE:  WordPerfect 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/862,730C
;      FILING DATE:  5/23/97
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1710 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      CELL TYPE:  canine peripheral blood macrophage
;      CELL LINE:  primary monocytes
;    IMMEDIATE SOURCE:
;      LIBRARY:  lambda gt11 cDNA
;      CLONE:  Canine IL-1ra
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1 to 1710
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  open reading frame
;      LOCATION:  60 to 587
;      OTHER INFORMATION:
US-08-862-730C-1


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-862-730C-1    ..

 Align seg 1/1  to: US-08-862-730C-1  from: 1  to: 1710

      31 SerProIleLysThrProSer 37
         ||||||||||||||||||||
    1358 TCTCCCATCAAAACTCCCAGC 1378

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-202-056-4

seq_documentation_block:
; Sequence 4, Application US/08202056
; Patent No. 5440021
;  GENERAL INFORMATION:
;    APPLICANT:  Chuntharapai, Anan
;    APPLICANT:  Hebert, Caroline
;    APPLICANT:  Kim, Kyung Jin
;    APPLICANT:  Lee, James
;    TITLE OF INVENTION:  Antibodies to Human IL-8 Type B Receptor
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/202,056
;       FILING DATE:  25-FEB-1994
;       CLASSIFICATION:  436
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/677211
;       FILING DATE:  29-MAR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Love, Richard B.
;       REGISTRATION NUMBER:  34,659
;       REFERENCE/DOCKET NUMBER:  706P3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-5530
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1737 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-202-056-4


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-202-056-4    ..

 Align seg 1/1  to: US-08-202-056-4  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         |||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-076-093A-3

seq_documentation_block:
; Sequence 3, Application US/08076093A
; Patent No. 5543503
;  GENERAL INFORMATION:
;     APPLICANT:  Chuntharapai, Anan
;     APPLICANT:  Lee, James
;     APPLICANT:  Hebert, Caroline
;     APPLICANT:  Jin Kim, K.
;     TITLE OF INVENTION:  Antibodies to Human PF4A Receptors
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/076,093A
;       FILING DATE:  11-Jun-1993
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/810782
;       FILING DATE:  19-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/677211
;       FILING DATE:  29-MAR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Love, Richard B
;       REGISTRATION NUMBER:  34,659
;       REFERENCE/DOCKET NUMBER:  706P2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-5530
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1737 nucleotides
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-076-093A-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-076-093A-3    ..

 Align seg 1/1  to: US-08-076-093A-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         |||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-701-265-3

seq_documentation_block:
; Sequence 3, Application US/08701265
; Patent No. 5776457
;  GENERAL INFORMATION:
;     APPLICANT:  Chuntharapai, Anan
;     APPLICANT:  Lee, James
;     APPLICANT:  Hebert, Caroline
;     APPLICANT:  Jin Kim, K.
;     TITLE OF INVENTION:  Antibodies to Human PF4A Receptors
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/701,265
;       FILING DATE:  22-AUG-1996
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/076093
;       FILING DATE:  11-Jun-1993
;       APPLICATION NUMBER:  07/810782
;       FILING DATE:  19-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/677211
;       FILING DATE:  29-MAR-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Love, Richard B
;       REGISTRATION NUMBER:  34,659
;       REFERENCE/DOCKET NUMBER:  706P2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-5530
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
```

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 nucleotides
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-701-265-3


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-701-265-3    ..

 Align seg 1/1  to: US-08-701-265-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         ||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-284-586-3

seq_documentation_block:
; Sequence 3, Application US/08284586
; Patent No. 5840856
;  GENERAL INFORMATION:
;    APPLICANT:  Chuntharapai, Anan
;    APPLICANT:  Lee, James
;    APPLICANT:  Hebert, Caroline
;    APPLICANT:  Jin Kim, K.
;    TITLE OF INVENTION:  Antibodies to Human PF4A Receptors
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/284,586
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,093A
;      FILING DATE:  11-Jun-1993
;      APPLICATION NUMBER:  07/810782
;      FILING DATE:  19-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/677211
;      FILING DATE:  29-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Love, Richard B
;      REGISTRATION NUMBER:  34,659
;      REFERENCE/DOCKET NUMBER:  706P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5530
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 nucleotides
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-284-586-3


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-284-586-3    ..

 Align seg 1/1  to: US-08-284-586-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         ||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-805-478-3

seq_documentation_block:
; Sequence 3, Application US/08805478
; Patent No. 5874543
;  GENERAL INFORMATION:
;    APPLICANT:  Chuntharapai, Anan
;    APPLICANT:  Lee, James
;    APPLICANT:  Hebert, Caroline
;    APPLICANT:  Jin Kim, K.
;    TITLE OF INVENTION:  ANTIBODIES TO PF4A RECEPTOR
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/805,478
;      FILING DATE:  25-Feb-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/284586
;      FILING DATE:  10-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076093
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/810782
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Love, Richard B.
;      REGISTRATION NUMBER:  34,659
;      REFERENCE/DOCKET NUMBER:  P0706P2P1C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5530
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-805-478-3


alignment_scores:
            Quality:    7.00            Length:      7
```

```
            Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-805-478-3   ..

 Align seg 1/1  to: US-08-805-478-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         ||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-802-627A-3

seq_documentation_block:
; Sequence 3, Application US/08802627A
; Patent No. 5892017
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, James
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING PF4A RECEPTOR
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/802,627A
;      FILING DATE:  19-Feb-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/284586
;      FILING DATE:  10-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076093
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/810782
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Love, Richard B.
;      REGISTRATION NUMBER:  34,659
;      REFERENCE/DOCKET NUMBER:  P0706P2P1D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5530
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-802-627A-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-802-627A-3   ..

 Align seg 1/1  to: US-08-802-627A-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         ||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-238-3

seq_documentation_block:
; Sequence 3, Application US/08801238
; Patent No. 5919896
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, James
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  PF4A RECEPTOR
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,238
;      FILING DATE:  19-Feb-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/284586
;      FILING DATE:  10-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076093
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/810782
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Love, Richard B.
;      REGISTRATION NUMBER:  34,659
;      REFERENCE/DOCKET NUMBER:  P0706P2P1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5530
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-801-238-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-801-238-3   ..

 Align seg 1/1  to: US-08-801-238-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         ||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-228-3
```

```
seq_documentation_block:
; Sequence 3, Application US/08801228
; Patent No. 5922541
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, James
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  METHODS FOR DETECTION AND AMPLIFICATION OF
;    TITLE OF INVENTION:  PF4A RECEPTOR NUCLEIC ACID
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,228
;      FILING DATE:  19-Feb-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/284586
;      FILING DATE:  10-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076093
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/810782
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Love, Richard B.
;      REGISTRATION NUMBER:  34,659
;      REFERENCE/DOCKET NUMBER:  P0706P2P1D3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5530
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-801-228-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-801-228-3    ..

 Align seg 1/1  to: US-08-801-228-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         ||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-104-296-3

seq_documentation_block:
; Sequence 3, Application US/09104296
; Patent No. 6087475
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, James
;    APPLICANT:  Wood, WIlliam I.
;    TITLE OF INVENTION:  PF4A Receptors
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/104,296
;      FILING DATE:  24-June-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/701265
;      FILING DATE:  22-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/664228
;      FILING DATE:  06-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076093
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/810782
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Love, Richard B.
;      REGISTRATION NUMBER:  34,659
;      REFERENCE/DOCKET NUMBER:  P0706P2C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5530
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-09-104-296-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-104-296-3    ..

 Align seg 1/1  to: US-09-104-296-3  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         ||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-06380-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9406380
;  GENERAL INFORMATION:
;    APPLICANT:  Chuntharapai, Anan
;    APPLICANT:  Lee, James
;    APPLICANT:  Hebert, Caroline
;    APPLICANT:  K. Jin Kim
;    APPLICANT:  Genentech, Inc.
;    TITLE OF INVENTION:  METHODS FOR TREATING INFLAMMATORY DISRODERS
;    NUMBER OF SEQUENCES:  3
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/06380
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/076093
;      FILING DATE:  11-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Love, Richard B.
;      REGISTRATION NUMBER:  34,659
;      REFERENCE/DOCKET NUMBER:  706P2P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5530
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1737 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US94-06380-2


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x PCT-US94-06380-2    ..

 Align seg 1/1  to: PCT-US94-06380-2  from: 1  to: 1737

     160 GlnCysAlaGlyGlyAlaAla 166
         |||||||||||||||||||||
       7 CAGTGTGCTGGCGGCGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-148-680-1

seq_documentation_block:
; Sequence 1, Application US/09148680
; Patent No. 6255561
;  GENERAL INFORMATION:
;    APPLICANT:  Kossmann, Jens
;    APPLICANT:  Willmitzer, Lothar
;    APPLICANT:  Emmermann, Michael
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES CODING FOR DEBRANCHING
;    TITLE OF INVENTION:  ENZYMES FROM MAIZE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & NEAVE
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10020-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPA)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/148,680
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP97/01141
;      FILING DATE:  06-MAR-97
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 196 08 918.2
;      FILING DATE:  07-MAR-96
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley Jr., James F.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  GFB-7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 596-9000
;      TELEFAX:  (212) 596-9090
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1993 base pairs
;      TYPE:  nucleotide
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Zea mays
;      TISSUE TYPE:  Blattgewebe
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1675
US-09-148-680-1


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-148-680-1    ..

 Align seg 1/1  to: US-09-148-680-1  from: 1  to: 1993

     150 ArgSerValTyrArgAsnPro 156
         |||||||||||||||||||||
    1509 CGAAGTGTCTATAGAAATCCC 1529

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-522-6

seq_documentation_block:
; Sequence 6, Application US/08749522
; Patent No. 6096950
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal
;    TITLE OF INVENTION:  FIBER-SPECIFIC PROMOTERS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  WI
;      COUNTRY:  U.S.A.
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/749,522
```

```
;       FILING DATE:
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Baker, Jean C.
;       REGISTRATION NUMBER:  35,433
;       REFERENCE/DOCKET NUMBER:  670513.90244
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (414) 277-5709
;       TELEFAX:  (414) 271-3552
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2168 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-749-522-6


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-749-522-6/rev   ..

 Align seg 1/1  to reverse of: US-08-749-522-6  from: 1  to: 2168

     136 AspLeuLysSerArgGlyIle 142
         |||||||||||||||||||||
     338 GATTTAAAAAGTAGAGGGATT 318

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-131-648-4

seq_documentation_block:
; Sequence 4, Application US/09131648
; Patent No. 6168920
; GENERAL INFORMATION:
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Yue, Henry
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Patterson, Chandra
;  TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
;  FILE REFERENCE: PF-0576 US
;  CURRENT APPLICATION NUMBER: US/09/131,648
;  CURRENT FILING DATE: 1998-08-10
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 2290
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 2687731
US-09-131-648-4


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-131-648-4   ..

 Align seg 1/1  to: US-09-131-648-4  from: 1  to: 2290

     143 LeuProAlaAsnThrGlnIle 149
         |||||||||||||||||||||
     261 TTGCCAGCTAACACACAGATT 281

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-176-620A-5

seq_documentation_block:
; Sequence 5, Application US/08176620A
; Patent No. 5595904
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G.
;    APPLICANT:  Cobb, Melanie H.
;    APPLICANT:  Yancopoulos, George D.
;    APPLICANT:  Nye, Steven
;    APPLICANT:  Panayotatos, Nikos
;    TITLE OF INVENTION:  A Family of Map2 Protein Kinases
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/176,620A
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6526-123
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  303..2018
US-08-176-620A-5


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-176-620A-5   ..

 Align seg 1/1  to: US-08-176-620A-5  from: 1  to: 3671

       8 PheIleThrThrLeuIleSer 14
         |||||||||||||||||||||
    1495 TTCATAACAACTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-463-862-5

seq_documentation_block:
; Sequence 5, Application US/08463862
; Patent No. 5776751
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G. et al.
;    TITLE OF INVENTION:  A FAMILY OF MAP2 PROTEIN KINASES
;    NUMBER OF SEQUENCES:  11
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,862
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/701,544
;      FILING DATE:  16-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6526-049
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  303..2018
US-08-463-862-5


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-463-862-5    ..

 Align seg 1/1  to: US-08-463-862-5  from: 1  to: 3671

       8 PheIleThrThrLeuIleSer 14
         ||||||||||||||||||||
    1495 TTCATAACAACTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-461-985-5

seq_documentation_block:
; Sequence 5, Application US/08461985
; Patent No. 5872006
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G.
;    APPLICANT:  Cobb, Melanie H.
;    APPLICANT:  Yancopoulos, George D.
;    APPLICANT:  Nye, Steven
;    APPLICANT:  Panayotatos, Nikos
;    TITLE OF INVENTION:  A Family of Map2 Protein Kinases
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,985
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/176,620
;      FILING DATE:  03-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  6526-123
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  303..2018
US-08-461-985-5


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-461-985-5    ..

 Align seg 1/1  to: US-08-461-985-5  from: 1  to: 3671

       8 PheIleThrThrLeuIleSer 14
         ||||||||||||||||||||
    1495 TTCATAACAACTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-458-887-5

seq_documentation_block:
; Sequence 5, Application US/08458887
; Patent No. 5914261
;  GENERAL INFORMATION:
;    APPLICANT:  Boulton, Teri G. et al.
;    TITLE OF INVENTION:  A FAMILY OF MAP2 PROTEIN KINASES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,887
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  536
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/701,544
;       FILING DATE:  16-MAY-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Misrock, S. Leslie
;       REGISTRATION NUMBER:  18,872
;       REFERENCE/DOCKET NUMBER:  6526-049
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212 790-9090
;       TELEFAX:  212 869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3671 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  303..2018
US-08-458-887-5


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-458-887-5    ..

 Align seg 1/1  to: US-08-458-887-5  from: 1  to: 3671

       8 PheIleThrThrLeuIleSer 14
         ||||||||||||||||||||
    1495 TTCATAACAACTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5

seq_documentation_block:
; Sequence 5, Application US/08932787B
; Patent No. 6277963
; GENERAL INFORMATION:
;  APPLICANT:  Boulton et al.
;  TITLE OF INVENTION:  ANTIBODIES DIRECTED TOWARD EXTRACELLULAR SIGNAL-RELATED
;  TITLE OF INVENTION:  KINASES
;  FILE REFERENCE:  REG 430-A-1
;  CURRENT APPLICATION NUMBER: US/08/932,787B
;  CURRENT FILING DATE:  1997-09-18
;  PRIOR APPLICATION NUMBER:  08/469,547
;  PRIOR FILING DATE:  1995-06-06
;  PRIOR APPLICATION NUMBER:  08/178,488
;  PRIOR FILING DATE:  1994-01-07
;  PRIOR APPLICATION NUMBER:  07/701,544
;  PRIOR FILING DATE:  1991-05-16
;  PRIOR APPLICATION NUMBER:  07/532,004
;  PRIOR FILING DATE:  1990-06-01
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 3671
;   TYPE: DNA
;   ORGANISM: RAT
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (303)...(2018)
;   OTHER INFORMATION: ERK3 cDNA
US-08-932-787B-5


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-932-787B-5    ..

 Align seg 1/1  to: US-08-932-787B-5  from: 1  to: 3671

       8 PheIleThrThrLeuIleSer 14
         ||||||||||||||||||||
    1495 TTCATAACAACTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-012C-5

seq_documentation_block:
; Sequence 5, Application US/08932012C
; Patent No. 6297035
; GENERAL INFORMATION:
;  APPLICANT: Boulton et al.
;  TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
;  FILE REFERENCE: REG 430-Y-1
;  CURRENT APPLICATION NUMBER: US/08/932,012C
;  CURRENT FILING DATE:  1997-09-17
;  PRIOR APPLICATION NUMBER: 08/462,874
;  PRIOR FILING DATE: 1995-06-05
;  PRIOR APPLICATION NUMBER: 08/178,488
;  PRIOR FILING DATE: 1994-01-07
;  PRIOR APPLICATION NUMBER: 07/701,544
;  PRIOR FILING DATE: 1991-05-16
;  PRIOR APPLICATION NUMBER: 07/532,004
;  PRIOR FILING DATE: 1990-06-01
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 3671
;   TYPE: DNA
;   ORGANISM: RAT
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (303)...(2018)
;   OTHER INFORMATION: ERK3 cDNA
US-08-932-012C-5


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-932-012C-5    ..

 Align seg 1/1  to: US-08-932-012C-5  from: 1  to: 3671

       8 PheIleThrThrLeuIleSer 14
         ||||||||||||||||||||
    1495 TTCATAACAACTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-888-818C-5

seq_documentation_block:
; Sequence 5, Application US/08888818C
; Patent No. 6303358
; GENERAL INFORMATION:
;  APPLICANT:  Boulton et al.
;  TITLE OF INVENTION:  A FAMILY OF MAP2 PROTEIN KINASES
;  FILE REFERENCE:  REG 430-V-1
;  CURRENT APPLICATION NUMBER: US/08/888,818C
;  CURRENT FILING DATE:  1997-07-07
;  PRIOR APPLICATION NUMBER:  08/478,985
;  PRIOR FILING DATE:  1995-06-07
;  PRIOR APPLICATION NUMBER:  08/178,488
;  PRIOR FILING DATE:  1994-01-07
```

```
;  PRIOR APPLICATION NUMBER:  07/701,544
;  PRIOR FILING DATE:  1991-05-16
;  PRIOR APPLICATION NUMBER:  07/532,004
;  PRIOR FILING DATE:  1990-06-01
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 3671
;   TYPE: DNA
;   ORGANISM: RAT
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (303)...(2018)
;   OTHER INFORMATION: ERK3 cDNA
US-08-888-818C-5


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-888-818C-5    ..

 Align seg 1/1  to: US-08-888-818C-5  from: 1  to: 3671

       8 PheIleThrThrLeuIleSer 14
         |||||||||||||||||||||
    1495 TTCATAACAACTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5459251-1

seq_documentation_block:
;Patent No. 5459251
;     APPLICANT: Tsujimoto, Yoshide;Croce, Carlo A.
;     TITLE OF INVENTION: DNA MOLECULES HAVING HUMAN BCL-2 GENE
;SEQUENCES
;     NUMBER OF SEQUENCES: 4
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/228,704
;       FILING DATE: 18-APR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 994,941
;       FILING DATE: 23-DEC-1992
;       APPLICATION NUMBER: 663,010
;       FILING DATE: 19-MAR-1991
;       APPLICATION NUMBER: 883,687
;       FILING DATE: 09-JUL-1986
;SEQ ID NO:1:
;       LENGTH: 4825
5459251-1


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x 5459251-1    ..

 Align seg 1/1  to: 5459251-1  from: 1  to: 4825

      31 SerProIleLysThrProSer 37
         |||||||||||||||||||||
    2446 AGTCCCATCAAAACTCCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-030-096-7

seq_documentation_block:
; Sequence 7, Application US/08030096
; Patent No. 5426041
;  GENERAL INFORMATION:
;    APPLICANT:  Fabijanski, Steven F.
;    APPLICANT:  Arnison, Paul G.
;    TITLE OF INVENTION:  BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID
;    TITLE OF INVENTION:  SEED PRODUCTION
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/030,096
;      FILING DATE:  22-MAR-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/556,917
;      FILING DATE:  20-JUL-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/CA91/00255
;      FILING DATE:  22-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  33229/164/PIHI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(2158..3225, 3663..4046)
US-08-030-096-7


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-030-096-7/rev    ..

 Align seg 1/1  to reverse of: US-08-030-096-7  from: 1  to: 4977

      99 SerGlnLeuLeuThrThrAla 105
         ||||||||||||||||||||
     374 TCGCAGCTGTTGACCACTGCA 354

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-485A-19

seq_documentation_block:
; Sequence 19, Application US/08465485A
; Patent No. 5831066
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John
;    TITLE OF INVENTION:  Regulation of bcl-2 Gene Expression
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
```

```
;       ADDRESSEE:  P.C.
;       STREET:  1755 S. Jefferson Davis Hwy., Suite 400
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/465,485A
;       FILING DATE:  05-JUN-1995
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/124,256
;       FILING DATE:  20-SEP-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/840,716
;       FILING DATE:  21-FEB-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/288,692
;       FILING DATE:  22-DEC-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fortney, Andrew D.
;       REGISTRATION NUMBER:  34,600
;       REFERENCE/DOCKET NUMBER:  3335-070-55 CONT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (408) 436-2070
;       TELEFAX:  (408) 436-2075
;   INFORMATION FOR SEQ ID NO:  19:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5086 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-465-485A-19


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-465-485A-19    ..

 Align seg 1/1  to: US-08-465-485A-19  from: 1  to: 5086

      31 SerProIleLysThrProSer 37
         |||||||||||||||||||||
    2446 AGTCCCATCAAAACTCCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-365-486A-14

seq_documentation_block:
; Sequence 14, Application US/08365486A
; Patent No. 5834306
;  GENERAL INFORMATION:
;    APPLICANT:  Webster, Keith A.
;    APPLICANT:  Bishopric, Nanette H.
;    TITLE OF INVENTION:  Tissue Specific Hypoxia Regulated
;    TITLE OF INVENTION:  Therapeutic Constructs
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,486A
;      FILING DATE:  23-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8255-0018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  human bcl-2 cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1459..2178
US-08-365-486A-14


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-365-486A-14    ..

 Align seg 1/1  to: US-08-365-486A-14  from: 1  to: 5086

      31 SerProIleLysThrProSer 37
         |||||||||||||||||||||
    2446 AGTCCCATCAAAACTCCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-080-285-19

seq_documentation_block:
; Sequence 19, Application US/09080285
; Patent No. 6040181
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John
;    TITLE OF INVENTION:  Regulation of bcl-2 Gene Expression
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Hwy., Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/080,285
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/465,485
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/124,256
;      FILING DATE:  20-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/840,716
;      FILING DATE:  21-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/288,692
;      FILING DATE:  22-DEC-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fortney, Andrew D.
;      REGISTRATION NUMBER:  34,600
;      REFERENCE/DOCKET NUMBER:  3335-070-55 CONT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (408) 436-2070
;      TELEFAX:  (408) 436-2075
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-080-285-19


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-080-285-19    ..

 Align seg 1/1  to: US-09-080-285-19  from: 1  to: 5086

      31 SerProIleLysThrProSer 37
         |||||||||||||||||||||
    2446 AGTCCCATCAAAACTCCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-880-342-14

seq_documentation_block:
; Sequence 14, Application US/08880342
; Patent No. 6218179
;  GENERAL INFORMATION:
;    APPLICANT:  Webster, Keith A.
;    APPLICANT:  Bishopric, Nanette H.
;    APPLICANT:  Murphy, Brian
;    APPLICANT:  Laderoute, Keith R.
;    APPLICANT:  Green, Christopher J.
;    TITLE OF INVENTION:  Tissue Specific Hypoxia Regulated
;    TITLE OF INVENTION:  Therapeutic Constructs
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/880,342
;      FILING DATE:  23-JUN-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/IB95/00996
;      FILING DATE:  13-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/365,486
;      FILING DATE:  23-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8255-0018.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  human bcl-2 cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1459..2178
US-08-880-342-14


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-08-880-342-14    ..

 Align seg 1/1  to: US-08-880-342-14  from: 1  to: 5086

      31 SerProIleLysThrProSer 37
         |||||||||||||||||||||
    2446 AGTCCCATCAAAACTCCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-05651-4

seq_documentation_block:
; Sequence 4, Application PC/TUS9305651
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  A Gene Which Prevents Programmed Cell Death
;    NUMBER OF SEQUENCES:  5
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  diskette
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05651
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1459..2178
;      OTHER INFORMATION:  /product= "Bcl-2"
PCT-US93-05651-4


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
 US-09-674-779-2 x PCT-US93-05651-4   ..

 Align seg 1/1  to: PCT-US93-05651-4  from: 1  to: 5086

      31 SerProIleLysThrProSer 37
         ||||||||||||||||||||
    2446 AGTCCCATCAAAACTCCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-06251-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9306251
;  GENERAL INFORMATION:
;    APPLICANT:  Wickstrom, Eric and Rife, Jason P.
;    TITLE OF INVENTION:  Trivalent Synthesis of Oligonucleotides Containing
;    TITLE OF INVENTION:  Stereospecific Alkylphosphonates and Arylphosphonates
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06251
;      FILING DATE:  19930630
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8586
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516-742-4343
;      TELEFAX:  516-742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5086 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US93-06251-2


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x PCT-US93-06251-2   ..

 Align seg 1/1  to: PCT-US93-06251-2  from: 1  to: 5086

      31 SerProIleLysThrProSer 37
         ||||||||||||||||||||
    2446 AGTCCCATCAAAACTCCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-234-186-7

seq_documentation_block:
; Sequence 7, Application US/09234186
; Patent No. 6312947
; GENERAL INFORMATION:
;  APPLICANT: Horvitz, H. Robert
;  APPLICANT: Hengartner, Michael
;  TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF A
;  TITLE OF INVENTION: GENE WHICH PROTECTS CELLS FROM PROGRAMMED CELL DEATH AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: 01997/201005
;  CURRENT APPLICATION NUMBER: US/09/234,186
;  CURRENT FILING DATE: 1999-01-20
;  EARLIER APPLICATION NUMBER: 07/898,933
;  EARLIER FILING DATE: 1992-06-12
;  EARLIER APPLICATION NUMBER: 07/927,681
;  EARLIER FILING DATE: 1992-08-10
;  EARLIER APPLICATION NUMBER: 08/288,295
;  EARLIER FILING DATE: 1994-08-10
;  EARLIER APPLICATION NUMBER: 08/801,248
;  EARLIER FILING DATE: 1997-02-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 5086
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1459)...(2178)
US-09-234-186-7


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x US-09-234-186-7   ..

 Align seg 1/1  to: US-09-234-186-7  from: 1  to: 5094

      31 SerProIleLysThrProSer 37
         ||||||||||||||||||||
    2454 AGTCCCATCAAAACTCCGTCT 2474

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5506344-1

seq_documentation_block:
;Patent No. 5506344
;    APPLICANT: TSUJIMOTO, YOSHIHIDE;CROCE, CARLO A.
;    TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2GENE PRODUCT
;    NUMBER OF SEQUENCES: 5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/435,193
;      FILING DATE: 05-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 228,704
;      FILING DATE: 18-APR-1994
;      APPLICATION NUMBER: 994,941
;      FILING DATE: 23-DEC-1992
;      APPLICATION NUMBER: 663,010
;      FILING DATE: 19-MAR-1991
;      APPLICATION NUMBER: 883,687
;      FILING DATE: 09-JUL-1986
;SEQ ID NO:1:
;      LENGTH: 5104
5506344-1


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x 5506344-1   ..

 Align seg 1/1  to: 5506344-1  from: 1  to: 5104
```

```
   31 SerProIleLysThrProSer 37
      ||||||||||||||||||||
 2446 AGTCCCATCAAAACTCCGTCT 2466
```

```
OM of: US-09-674-779-2 to: EST:*   out_format : pfs

Date: Sep 16, 2002  7:24 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124143_26847/app_query.fasta_1.310
 -DB=EST -QFMT=fastap -SUFFIX=oli.rst -GAPOP=4.500 -GAPEXT=0.050
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=45
 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0
 -MAXLEN=2000000000 -USER=US09674779_@CGN1_1_2604 -NCPU=6
 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1623.650000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est2:BF353182       +    9.00   142.58   107.45    579  ! BF353182 PM1-HT0628-310800-009-
gb_est2:BF353196       +    9.00   142.58   107.45    579  ! BF353196 PM1-HT0629-290800-009-
gb_est2:BE344072       +    9.00   142.26   112.00    602  ! BE344072 EST409234 potato stolc
gb_gss:AZ374621        +    9.00   141.18   128.69    686  ! AZ374621 1M0127G01R Mouse 10kb
gb_gss:CNS020HD        -    9.00   140.22   145.51    770  ! AL175594 Tetraodon nigroviridis
gb_gss:CNS05M5R        +    9.00   138.25   187.45    977  ! AL343656 Tetraodon nigroviridis
gb_gss:CNS05P19        +    9.00   137.78   199.11   1034  ! AL347382 Tetraodon nigroviridi
gb_gss:CNS05FU5        +    9.00   137.25   212.86   1101  ! AL335462 Tetraodon nigroviridi
gb_gss:CNS02KVL        -    9.00   137.21   214.09   1107  ! AL202026 Tetraodon nigroviridi
gb_gss:BH233428        +    8.00   136.48   235.13    149  ! BH233428 1006173F09.2EL_x1 1006
gb_est2:BE708381       -    8.00   134.16   316.45    197  ! BE708381 MR0-HT0559-060600-030-
gb_est1:AV348302       -    8.00   132.12   411.19    252  ! AV348302 AV348302 RIKEN full-le
gb_est1:AA557886       -    8.00   132.05   414.66    254  ! AA557886 nl64f02.s1 NCI_CGAP_Pr
gb_est2:BG375333       +    8.00   131.25   459.95    280  ! BG375333 UI-R-CV1-bsz-b-12-0-UI
gb_est2:BI510695       +    8.00   130.56   502.00    304  ! BI510695 BB160003A20C06 Bee Bra
gb_est1:BB042909       +    8.00   130.54   503.76    305  ! BB042909 BB042909 RIKEN full-le
gb_est2:BG957563       -    8.00   130.46   509.03    308  ! BG957563 CM2-CT0662-020301-696-
gb_est2:BI509041       +    8.00   130.35   516.06    312  ! BI509041 BB170015B20G01 Bee Bra
gb_est1:AI613633       -    8.00   129.03   611.57    366  ! AI613633 vg30c10.y1 Soares_mamm
gb_gss:TA262B01Q       +    8.00   128.67   640.04    382  ! AL487669 T. brucei sheared genc
gb_est2:BF934922       -    8.00   128.44   659.66    393  ! BF934922 IL2-NT0201-231200-323-
gb_gss:AQ131150        -    8.00   128.23   677.53    403  ! AQ131150 HS_3036_A1_F06_MR CIT
gb_est1:AV544347       +    8.00   128.08   690.06    410  ! AV544347 AV544347 Arabidopsis t
gb_gss:AQ219969        +    8.00   127.90   706.18    419  ! AQ219969 HS_3251_B2_G07_MR CIT
gb_est2:BG913055       -    8.00   127.77   718.73    426  ! BG913055 602806433F1 NCI_CGAP_E
gb_est2:R40672         -    8.00   127.69   725.91    430  ! R40672 yf79e11.s1 Soares infant
gb_est2:BI513696       +    8.00   127.42   751.08    444  ! BI513696 BB160013B10D03 Bee Bra
gb_est2:BF562471       -    8.00   127.39   754.68    446  ! BF562471 UI-R-BS0-ans-e-07-0-UI
gb_gss:AQ600079        +    8.00   127.39   754.68    446  ! AQ600079 HS_5354_B2_G05_SP6E RP
gb_est1:AW215765       -    8.00   127.22   770.89    455  ! AW215765 up09g09.y1 NCI_CGAP_Lu
gb_est1:AA509926       -    8.00   127.20   772.69    456  ! AA509926 vg30c10.r1 Soares_mamm
gb_gss:B45889          -    8.00   127.17   776.29    458  ! B45889 HS-1062-B1-C12-MR.abi CI
gb_est1:AA874458       -    8.00   127.15   778.10    459  ! AA874458 vx80e06.r1 Soares_thym
gb_est2:BI514220       +    8.00   127.06   787.12    464  ! BI514220 BB160014B10H08 Bee Bra
gb_est2:BF565248       -    8.00   127.04   788.92    465  ! BF565248 UI-R-BO1-ajj-c-10-0-UI
gb_est2:BI513342       +    8.00   126.83   810.59    477  ! BI513342 BB160012B10B07 Bee Bra
gb_gss:BH105280        -    8.00   126.66   828.68    487  ! BH105280 RPCI-24-351K23.TJ RPCI
gb_est2:BI512001       +    8.00   126.62   832.30    489  ! BI512001 BB160007B20D07 Bee Bra
gb_est2:BI515132       +    8.00   126.18   881.27    516  ! BI515132 BB160017B10C08 Bee
gb_gss:B48301          +    8.00   125.99   903.08    528  ! B48301 RPCI11-6P4.TV RPCI-11
gb_est2:BI643242       -    8.00   125.91   912.18    533  ! BI643242 RS4_C01 Sugar beet
gb_est2:BI517071       +    8.00   125.88   915.82    535  ! BI517071 BB160024A20E01 Bee
gb_est1:AI995165       -    8.00   125.74   932.22    544  ! AI995165 701502426 A. thalia
gb_gss:AQ266185        +    8.00   125.22   996.14    579  ! AQ266185 CITBI-E1-2502L22.TF
gb_est2:BI515482       +    8.00   125.21   997.97    580  ! BI515482 BB160019B10B04 Bee

seq_name: gb_est2:BF353182

seq_documentation_block:
LOCUS       BF353182                579 bp    mRNA    linear   EST 22-NOV-2000
DEFINITION  PM1-HT0628-310800-009-e08 HT0628 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF353182
VERSION     BF353182.1  GI:11312256
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-HT0628-
            310800-009-e08&t3=2000-08-31&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 20
            High quality sequence stop: 46.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0628"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      190 a    131 c    137 g    121 t
ORIGIN


alignment_scores:
           Quality:    9.00              Length:      9
             Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BF353182   ..

 Align seg 1/1  to: BF353182  from: 1  to: 579
```

```
     85 AlaTyrLeuGlnSerArgLeuGlyAsn 93
        |||||||||||||||||||||||||||
    123 GCTTACCTACAAAGCCGGCTGGGCAAT 149

seq_name: gb_est2:BF353196

seq_documentation_block:
LOCUS       BF353196                579 bp    mRNA    linear   EST 22-NOV-2000
DEFINITION  PM1-HT0629-290800-009-e08 HT0629 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF353196
VERSION     BF353196.1  GI:11312270
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM1&t2=PM1-HT0629-
            290800-009-e08&t3=2000-08-29&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 20
            High quality sequence stop: 46.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0629"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      190 a    131 c    137 g    121 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BF353196   ..

 Align seg 1/1  to: BF353196  from: 1  to: 579

     85 AlaTyrLeuGlnSerArgLeuGlyAsn 93
        |||||||||||||||||||||||||||
    123 GCTTACCTACAAAGCCGGCTGGGCAAT 149


seq_name: gb_est2:BE344072

seq_documentation_block:
LOCUS       BE344072                602 bp    mRNA    linear   EST 17-JUL-2000
DEFINITION  EST409234 potato stolon, Cornell University Solanum tuberosum cDNA
            clone cSTA28A10, mRNA sequence.
ACCESSION   BE344072
VERSION     BE344072.1  GI:9253604
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
            Holt,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan
            ,B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker
            ,B.
  TITLE     Generation of ESTs from potato swelling stolons
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Research Genetics, Libraries Division
            Tel: 1-800-711-6195
            Email: cdna@resgen.com
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Solanum tuberosum"
                     /cultivar="Bintje"
                     /db_xref="taxon:4113"
                     /clone="cSTA28A10"
                     /clone_lib="potato stolon, Cornell University"
                     /tissue_type="axillary buds of stem explants, swelling
                     stolons"
                     /dev_stage="1 to 3 days"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; RNA was supplied by Christian Bachem & Beatrix
                     Horvath(Laboratory of Plant Breeding, Dept. of Plant
                     Sciences,  Wageningen University, The Netherlands). Total
                     RNA was isolated from developing axillary buds of potato
                     nodal stem cuttings  cultured on medium for the
                     introduction of tuber formation as  described in Bachem et
                     al. (Plant Journal 1996). Tissue samples were  taken of
                     stages corresponding to growing stolons and the early
                     stages  of tuber formation."
BASE COUNT      183 a    117 c    127 g    175 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BE344072   ..

 Align seg 1/1  to: BE344072  from: 1  to: 602

     99 SerGlnLeuLeuThrThrAlaArgSer 107
        |||||||||||||||||||||||||||
    497 TCCCAACTGCTAACAACAGCGCGTTCT 523

seq_name: gb_gss:AZ374621

seq_documentation_block:
LOCUS       AZ374621                686 bp    DNA     linear   GSS 02-OCT-2000
DEFINITION  1M0127G01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0127G01 R, DNA sequence.
ACCESSION   AZ374621
VERSION     AZ374621.1  GI:10488321
KEYWORDS    GSS.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0127  row: G  column: 01
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0127G01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      174 a    168 c    110 g    234 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AZ374621   ..

 Align seg 1/1  to: AZ374621  from: 1  to: 686

       7 TyrPheIleThrThrLeuIleSerSer 15
         |||||||||||||||||||||||||||
     178 TACTTCATAACTACATTAATCAGTTCA 204

seq_name: gb_gss:CNS020HD

seq_documentation_block:
LOCUS       CNS020HD                 770 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222024 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175594
VERSION     AL175594.1  GI:7813651
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 770)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 770)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 770)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .770
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="222024"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222BH12LP1-end : T7"
BASE COUNT      233 a    140 c    179 g    213 t      5 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x CNS020HD/rev   ..

 Align seg 1/1  to reverse of: CNS020HD  from: 1 . to: 770

     116 TyrGlnLeuProProGluHisLeuTrp 124
         |||||||||||||||||||||||||||
     581 TACCAGCTGCCTCCAGAACACTTGTGG 555

seq_name: gb_gss:CNS05M5R

seq_documentation_block:
LOCUS       CNS05M5R                 977 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            029008 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL343656
VERSION     AL343656.1  GI:8237426
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
           Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
           Tetraodontidae; Tetraodon.
REFERENCE  1  (bases 1 to 977)
  AUTHORS  Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
           Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
           Weissenbach,J.
  TITLE    Charaterization and repeat analysis of the compact genome of the
           freshwater pufferfish Tetraodon nigroviridis
  JOURNAL  Unpublished
REFERENCE  2  (bases 1 to 977)
  AUTHORS  Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
           Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
           Saurin,W. and Weissenbach,J.
  TITLE    Human gene number estimate provided by genome wide analysis using
           Tetraodon nigroviridis DNA sequence
  JOURNAL  Unpublished
REFERENCE  3  (bases 1 to 977)
  AUTHORS  Genoscope.
  TITLE    Direct Submission
  JOURNAL  Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT    This sequence is a single read and was generated as part of a large
           scale clone-end sequencing project of the Tetraodon nigroviridis
           genome. For more information, please take a look at
           http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .977
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="029008"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA029BH04A1~end : T3"
BASE COUNT      272 a    263 c    198 g    217 t     27 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x CNS05M5R   ..

 Align seg 1/1  to: CNS05M5R  from: 1  to: 977

     116 TyrGlnLeuProProGluHisLeuTrp 124
         ||||||||||||||||||||||||||||
     714 TACCAGCTGCCTCCAGAACACTTGTGG 740

seq_name: gb_gss:CNS05P19

seq_documentation_block:
LOCUS       CNS05P19                1034 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            035C15 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL347382
VERSION     AL347382.1  GI:8241152
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1  (bases 1 to 1034)
  AUTHORS  Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
           Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
           Weissenbach,J.
  TITLE    Charaterization and repeat analysis of the compact genome of the
           freshwater pufferfish Tetraodon nigroviridis
  JOURNAL  Unpublished
REFERENCE  2  (bases 1 to 1034)
  AUTHORS  Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
           Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
           Saurin,W. and Weissenbach,J.
  TITLE    Human gene number estimate provided by genome wide analysis using
           Tetraodon nigroviridis DNA sequence
  JOURNAL  Unpublished
REFERENCE  3  (bases 1 to 1034)
  AUTHORS  Genoscope.
  TITLE    Direct Submission
  JOURNAL  Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT    This sequence is a single read and was generated as part of a large
           scale clone-end sequencing project of the Tetraodon nigroviridis
           genome. For more information, please take a look at
           http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1034
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="035C15"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA035AB08A1~end : T3"
BASE COUNT      285 a    284 c    213 g    230 t     22 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x CNS05P19   ..

 Align seg 1/1  to: CNS05P19  from: 1  to: 1034

     116 TyrGlnLeuProProGluHisLeuTrp 124
         ||||||||||||||||||||||||||||
     718 TACCAGCTGCCTCCAGAACACTTGTGG 744

seq_name: gb_gss:CNS05FU5

seq_documentation_block:
LOCUS       CNS05FU5                1101 bp    DNA     linear   GSS 26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            005M17 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL335462
VERSION     AL335462.1  GI:8229220
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1  (bases 1 to 1101)
  AUTHORS  Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
           Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
           Weissenbach,J.
  TITLE    Charaterization and repeat analysis of the compact genome of the
           freshwater pufferfish Tetraodon nigroviridis
  JOURNAL  Unpublished
REFERENCE  2  (bases 1 to 1101)
  AUTHORS  Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
           Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
           Saurin,W. and Weissenbach,J.
  TITLE    Human gene number estimate provided by genome wide analysis using
           Tetraodon nigroviridis DNA sequence
  JOURNAL  Unpublished
REFERENCE  3  (bases 1 to 1101)
  AUTHORS  Genoscope.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="005M17"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA005AG09C1~end : T7"
BASE COUNT      298 a    311 c    221 g    264 t      7 others
ORIGIN


alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x CNS05FU5    ..

 Align seg 1/1  to: CNS05FU5  from: 1  to: 1101

     116 TyrGlnLeuProProGluHisLeuTrp 124
         ||||||||||||||||||||||||||||
     729 TACCAGCTGCCTCCAGAACACTTGTGG 755

seq_name: gb_gss:CNS02KVL

seq_documentation_block:
LOCUS       CNS02KVL                1107 bp    DNA     linear   GSS 14-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            146C02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL202026
VERSION     AL202026.1  GI:7860371
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1107)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1107)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1107)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1107
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="146C02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG146BB01LP1~end : T7"
BASE COUNT      300 a    221 c    270 g    310 t      6 others
ORIGIN


alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x CNS02KVL/rev    ..

 Align seg 1/1  to reverse of: CNS02KVL  from: 1  to: 1107

     116 TyrGlnLeuProProGluHisLeuTrp 124
         ||||||||||||||||||||||||||||
     481 TACCAGCTGCCTCCAGAACACTTGTGG 455

seq_name: gb_gss:BH233428

seq_documentation_block:
LOCUS       BH233428                149 bp    DNA     linear   GSS 08-NOV-2001
DEFINITION  1006173F09.2EL_x1 1006 - RescueMu Grid G Zea mays genomic, DNA
            sequence.
ACCESSION   BH233428
VERSION     BH233428.1  GI:16839654
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 149)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1006173  row: 5
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .149
                     /organism="Zea mays"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /clone_lib="1006 - RescueMu Grid G"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid G was grown at Stanford in 2000. DNA was
                     extracted from leaf punches, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
BASE COUNT       32 a     36 c     21 g     60 t
ORIGIN
```

```
alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BH233428   ..

 Align seg 1/1  to: BH233428  from: 1  to: 149

     170 HisLeuThrAsnSerAlaIleAsp 177
         |||||||||||||||||||||||
      28 CACCTAACCAACTCGGCTATAGAC 51

seq_name: gb_est2:BE708381

seq_documentation_block:
LOCUS       BE708381                197 bp    mRNA    linear   EST 12-SEP-2000
DEFINITION  MR0-HT0559-060600-030-c10 HT0559 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE708381
VERSION     BE708381.1  GI:10096646
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR0-HT0559-060
            600-030-c10&t3=2000-06-06&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 22
            High quality sequence stop: 137.
FEATURES             Location/Qualifiers
     source          1. .197
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0559"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       55 a     46 c     47 g     49 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BE708381/rev   ..

 Align seg 1/1  to reverse of: BE708381  from: 1  to: 197

      58 SerHisValGlyLeuGlnAlaHis 65
         |||||||||||||||||||||||
     105 TCACACGTTGGACTGCAAGCACAC 82

seq_name: gb_est1:AV348302

seq_documentation_block:
LOCUS       AV348302                252 bp    mRNA    linear   EST 12-NOV-1999
DEFINITION  AV348302 RIKEN full-length enriched, adult male olfactory bulb Mus
            musculus cDNA clone 6430704H12 3', mRNA sequence.
ACCESSION   AV348302
VERSION     AV348302.1  GI:6389361
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .252
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430704H12"
                     /clone_lib="RIKEN full-length enriched, adult male
                     olfactory bulb"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
```

/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 67 a 48 c 71 g 66 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AV348302/rev ..

Align seg 1/1 to reverse of: AV348302 from: 1 to: 252

219 AlaIleHisLeuAspThrGlnGly 226
||||||||||||||||||||||||
134 GCTATACACCTGGATACGCAGGGC 111

seq_name: gb_est1:AA557886

seq_documentation_block:
LOCUS AA557886 254 bp mRNA linear EST 09-SEP-1997
DEFINITION n164f02.s1 NCI_CGAP_Pr4.1 Homo sapiens cDNA clone IMAGE:1045467, mRNA sequence.
ACCESSION AA557886
VERSION AA557886.1 GI:2328363
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 254)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 359 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES Location/Qualifiers
source 1. .254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1045467"
/clone_lib="NCI_CGAP_Pr4.1"
/sex="male"
/tissue_type="prostatic intraepithelial neoplasia - high grade"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from prostatic intraepithelial neoplasia (high-grade), cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. cDNA Library Preparation: David B. Krizman, Ph.D. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center"

BASE COUNT 76 a 63 c 51 g 64 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AA557886/rev ..

Align seg 1/1 to reverse of: AA557886 from: 1 to: 254

211 GlyLeuGlyLeuTyrAlaThrGly 218
||||||||||||||||||||||||
38 GGTTTGGGGCTTTATGCCACTGGG 15

seq_name: gb_est2:BG375333

seq_documentation_block:
LOCUS BG375333 280 bp mRNA linear EST 12-MAR-2001
DEFINITION UI-R-CV1-bsz-b-12-0-UI.s1 UI-R-CV1 Rattus norvegicus cDNA clone UI-R-CV1-bsz-b-12-0-UI 3', mRNA sequence.
ACCESSION BG375333
VERSION BG375333.1 GI:13299805
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 280)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized rat eye library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-35, >AT_rich#Low_complexity
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .280
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"

```
                     /db_xref="taxon:10116"
                     /clone="UI-R-CV1-bsz-b-12-0-UI"
                     /clone_lib="UI-R-CV1"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CV1
                     library is a normalized library constructed from rat eye
                     tissue. For a detailed description of the library from
                     which this clone was derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=UI-R-CV1
                     TAG_TISSUE=rat eye
                     TAG_SEQ=CAGCC"
BASE COUNT       75 a     56 c     52 g     97 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BG375333   ..

 Align seg 1/1  to: BG375333  from: 1  to: 280

      14 SerSerMetLeuValAlaCysSer 21
         |||||||||||||||||||||||||
     252 AGCTCGATGTTAGTTGCGTGCAGC 275

seq_name: gb_est2:BI510695

seq_documentation_block:
LOCUS       BI510695                304 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160003A20C06 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160003A20C06 5', mRNA sequence.
ACCESSION   BI510695
VERSION     BI510695.1  GI:15361069
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 304)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 304   Std Error: 0.00
            Plate: BB160003A20  row: C  column: 06
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 304.
FEATURES             Location/Qualifiers
     source          1. .304
                     /organism="Apis mellifera"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB160003A20C06"
                     /clone_lib="Bee Brain Normalized Library, BB16"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoRl;
                     Site_2: Notl; The BB16 library was contributed by the
                     Soares laboratory and it was  constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G.  and Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared from dissected brains of adult worker bees of
                     various ages  and various behavioral groups. "
BASE COUNT       82 a     60 c     72 g     90 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI510695   ..

 Align seg 1/1  to: BI510695  from: 1  to: 304

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||||
      75 CAGAATTTCGGATTAGGCCTTTAT 98

seq_name: gb_est1:BB042909

seq_documentation_block:
LOCUS       BB042909                305 bp    mRNA    linear   EST 23-JUN-2000
DEFINITION  BB042909 RIKEN full-length enriched, 13 days embryo male testis Mus
            musculus cDNA clone 6030468A09 3', mRNA sequence.
ACCESSION   BB042909
VERSION     BB042909.1  GI:8449295
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 305)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
```

```
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES                Location/Qualifiers
     source             1. .305
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="6030468A09"
                        /clone_lib="RIKEN full-length enriched, 13 days embryo
                        male testis"
                        /sex="male"
                        /tissue_type="testis"
                        /dev_stage="13 days embryo"
                        /lab_host="DH10B"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. cDNA went through one round of normalization
                        to Rot = 5.0 and subtraction to Rot = 100.0. Second strand
                        cDNA was prepared with the primer adapter of sequence [5'
                        GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                        was cloned into the XhoI and BamHI sites. Vector: a
                        modified pBluescript KS(+) after bulk excision from Lambda
                        FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT       63 a     82 c     50 g    110 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BB042909    ..

 Align seg 1/1  to: BB042909  from: 1  to: 305

     137 LeuLysSerArgGlyIleLeuPro 144
         |||||||||||||||||||||||||
      78 CTCAAGTCCAGAGGAATCCTGCCT 101

seq_name: gb_est2:BG957563

seq_documentation_block:
LOCUS       BG957563                 308 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  CM2-CT0662-020301-696-d09 CT0662 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG957563
VERSION     BG957563.1  GI:14375734
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-CT0662-
            020301-696-d09&t3=2001-03-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 308.
FEATURES                Location/Qualifiers
     source             1. .308
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="CT0662"
                        /dev_stage="Adult"
                        /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                        SmaI; A mini-library was made by cloning products derived
                        from  ORESTES PCR (U.S. Letters Patent application No. 196
                        ,716 -  Ludwig Institute for Cancer Research) profiles
                        into  the pUC 18 vector. Reverse transcription of tissue
                        mRNA and cDNA amplification were performed under low
                        stringency conditions."
BASE COUNT       64 a    105 c     83 g     56 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BG957563/rev    ..

 Align seg 1/1  to reverse of: BG957563  from: 1  to: 308

     101 LeuLeuThrThrAlaArgSerTrp 108
         ||||||||||||||||||||||||
     250 CTTCTAACAACTGCGCGTTCATGG 227

seq_name: gb_est2:BI509041

seq_documentation_block:
LOCUS       BI509041                 312 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB170015B20G01 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170015B20G01 5', mRNA sequence.
ACCESSION   BI509041
VERSION     BI509041.1  GI:15359415
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 312)
```

```
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 312   Std Error: 0.00
            Plate: BB170015B20  row: G  column: 01
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 312.
FEATURES              Location/Qualifiers
     source           1. .312
                      /organism="Apis mellifera"
                      /strain="mixed strains of European bees, predominantly
                      A.m. ligustica"
                      /db_xref="taxon:7460"
                      /clone="BB170015B20G01"
                      /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
                      /sex="female"
                      /tissue_type="brain"
                      /dev_stage="adult worker honey bee"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                      Site_2: Not1; This BB17 cDNA library was generated by
                      subtraction of the BB16  library with 4000 previously
                      sequenced clones. The BB16 library was  contributed by the
                      Soares laboratory and it was constructed and  normalized
                      as described by Bonaldo, M.F., Lennon, G. and  Soares,
                      M.B. (1996), Genome Research 6(9): 791-806. RNA was
                      prepared  from dissected brains of adult worker bees of
                      various ages and  various behavioral groups. "
BASE COUNT       83 a     60 c     73 g     96 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI509041   ..

 Align seg 1/1  to: BI509041  from: 1  to: 312

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||
      73 CAGAATTTCGGATTAGGCCTTTAT 96

seq_name: gb_est1:AI613633

seq_documentation_block:
LOCUS       AI613633               366 bp    mRNA    linear   EST 21-APR-1999
DEFINITION  vg30c10.y1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:862866 5', mRNA sequence.
ACCESSION   AI613633
VERSION     AI613633.1  GI:4622800
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:506954
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 365.
FEATURES              Location/Qualifiers
     source           1. .366
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:862866"
                      /clone_lib="Soares_mammary_gland_NbMMG"
                      /sex="male"
                      /tissue_type="mammary gland"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
                      /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                      ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                      RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                      primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
BASE COUNT       92 a     114 c     108 g     52 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AI613633/rev   ..

 Align seg 1/1  to reverse of: AI613633  from: 1  to: 366

     101 LeuLeuThrThrAlaArgSerTrp 108
         |||||||||||||||||||||||
     190 CTTCTAACAACTGCGCGTTCATGG 167

seq_name: gb_gss:TA262B01Q

seq_documentation_block:
LOCUS       TA262B01Q              382 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 262b01, reverse sequence,
            genomic survey sequence.
ACCESSION   AL487669
VERSION     AL487669.1  GI:11850705
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
```

Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh1@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution ( 4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

| FEATURES | Location/Qualifiers |
|---|---|
| source | 1. .382 |
| | /organism="Trypanosoma brucei" |
| | /strain="TREU927" |
| | /db_xref="taxon:5691" |
| | /clone="262b01" |

BASE COUNT 99 a 79 c 61 g 142 t 1 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x TA262B01Q ..

Align seg 1/1 to: TA262B01Q from: 1 to: 382

123 LeuTrpGlyGlnIleValProThr 130
||||||||||||||||||||||||
53 TTGTGGGGTCAAATAGTCCCTACC 76

seq_name: gb_est2:BF934922

seq_documentation_block:
LOCUS BF934922 393 bp mRNA linear EST 22-JAN-2001
DEFINITION IL2-NT0201-231200-323-H06 NT0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF934922
VERSION BF934922.1 GI:12352246
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 393)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL2&t2=IL2-NT0201-231200-323-H06&t3=2000-12-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 392.

| FEATURES | Location/Qualifiers |
|---|---|
| source | 1. .393 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /clone_lib="NT0201" |
| | /dev_stage="Adult" |
| | /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." |

BASE COUNT 121 a 69 c 82 g 120 t 1 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x BF934922/rev ..

Align seg 1/1 to reverse of: BF934922 from: 1 to: 393

101 LeuLeuThrThrAlaArgSerTrp 108
||||||||||||||||||||||||
331 CTTTTGACCACTGCTCGGTCCTGG 308

seq_name: gb_gss:AQ131150

seq_documentation_block:
LOCUS AQ131150 403 bp DNA linear GSS 23-SEP-1998
DEFINITION HS_3036_A1_F06_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3036 Col=11 Row=K, DNA sequence.
ACCESSION AQ131150
VERSION AQ131150.1 GI:3508316
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 403)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3036 row: K column: 11
Class: BAC ends
High quality sequence stop: 403.

| FEATURES | Location/Qualifiers |
|---|---|
| source | 1. .403 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |

```
                    /clone="Plate=3036 Col=11 Row=K"
                    /clone_lib="CIT Approved Human Genomic Sperm Library D"
                    /sex="male"
                    /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                    E-Coli DH10B"
BASE COUNT       95 a    112 c     53 g    143 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AQ131150/rev   ..

 Align seg 1/1  to reverse of: AQ131150  from: 1  to: 403

     137 LeuLysSerArgGlyIleLeuPro 144
         ||||||||||||||||||||||||
     322 CTCAAAAGCAGGGGGATTTTGCCA 299

seq_name: gb_est1:AV544347

seq_documentation_block:
LOCUS       AV544347               410 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  AV544347 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ40g10F 3', mRNA sequence.
ACCESSION   AV544347
VERSION     AV544347.1  GI:8715761
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZ40g10F"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /tissue_type="roots"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      128 a    106 c     56 g    120 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AV544347   ..

 Align seg 1/1  to: AV544347  from: 1  to: 410

     144 ProAlaAsnThrGlnIleArgSer 151
         |||||||||||||||||||||||
     152 CCGGCCAATACACAGATCCGATCC 175

seq_name: gb_gss:AQ219969

seq_documentation_block:
LOCUS       AQ219969               419 bp    DNA     linear   GSS 19-SEP-1998
DEFINITION  HS_3251_B2_G07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3251 Col=14 Row=N, DNA sequence.
ACCESSION   AQ219969
VERSION     AQ219969.1  GI:3633582
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3251  row: N  column: 14
            Class: BAC ends
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3251 Col=14 Row=N"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      117 a     89 c     80 g    126 t      7 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AQ219969   ..

 Align seg 1/1  to: AQ219969  from: 1  to: 419

     230 TrpGlyAlaGlnPheSerGluThr 237
         |||||||||||||||||||||||
     206 TGGGGAGCTCAGTTCTCAGAAACT 229

seq_name: gb_est2:BG913055

seq_documentation_block:
LOCUS       BG913055               426 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  602806433F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4938513
            5', mRNA sequence.
ACCESSION   BG913055
VERSION     BG913055.1  GI:14293531
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10874  row: a  column: 10
            High quality sequence stop: 426.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4938513"
                     /clone_lib="NCI_CGAP_Brn67"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT       97 a    106 c     82 g    141 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BG913055/rev   ..

 Align seg 1/1  to reverse of: BG913055  from: 1  to: 426

     167 MetSerLysHisLeuThrAsnSer 174
         ||||||||||||||||||||||||
     253 ATGAGCAAACATCTGACAAATTCA 230

seq_name: gb_est2:R40672

seq_documentation_block:
LOCUS       R40672                   430 bp    mRNA    linear   EST 22-MAY-1995
DEFINITION  yf79e11.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:28343 3', mRNA sequence.
ACCESSION   R40672
VERSION     R40672.1  GI:821010
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 5, 1995 this sequence version replaced gi:798288.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1430
            High quality sequence stops: 251 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1430   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 251.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Homo sapiens"
                     /db_xref="GDB:400690"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:28343"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      115 a     81 c     84 g    146 t      4 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x R40672/rev   ..

 Align seg 1/1  to reverse of: R40672  from: 1  to: 430

      24 IleProThrAsnProGlnValSer 31
         |||||||||||||||||||||||
     374 ATACCCACNAATCCACAGGTGAGT 351

seq_name: gb_est2:BI513696

seq_documentation_block:
LOCUS       BI513696                 444 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160013B10D03 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160013B10D03 5', mRNA sequence.
ACCESSION   BI513696
VERSION     BI513696.1  GI:15364070
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
```

```
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 444   Std Error: 0.00
            Plate: BB160013B10  row: D  column: 03
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 444.
FEATURES              Location/Qualifiers
     source           1. .444
                      /organism="Apis mellifera"
                      /strain="mixed strains of European bees, predominantly
                      A.m. ligustica"
                      /db_xref="taxon:7460"
                      /clone="BB160013B10D03"
                      /clone_lib="Bee Brain Normalized Library, BB16"
                      /sex="female"
                      /tissue_type="brain"
                      /dev_stage="adult worker honey bee"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                      Site_2: Not1; The BB16 library was contributed by the
                      Soares laboratory and it was  constructed and normalized
                      as described by Bonaldo, M.F., Lennon, G.  and Soares,
                      M.B. (1996), Genome Research 6(9): 791-806. RNA was
                      prepared from dissected brains of adult worker bees of
                      various ages  and various behavioral groups. "
BASE COUNT      149 a     69 c     95 g    131 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI513696   ..

 Align seg 1/1  to: BI513696  from: 1  to: 444

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||
      63 CAAAATTTTGGTTTGGGATTATAT 86

seq_name: gb_est2:BF562471

seq_documentation_block:
LOCUS       BF562471                446 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  UI-R-BS0-ans-e-07-0-UI.r1 UI-R-BS0 Rattus norvegicus cDNA clone
            UI-R-BS0-ans-e-07-0-UI 5', mRNA sequence.
ACCESSION   BF562471
VERSION     BF562471.1  GI:11672249
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1801788
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .446
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-BS0-ans-e-07-0-UI"
                      /clone_lib="UI-R-BS0"
                      /dev_stage="embryonic 13 dpc"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS0
                      library is derived from 13 dpc whole embryo tissue. For a
                      detailed description of the library from which this clone
                      was derived, please visit our web site at
                      ratest.eng.uiowa.edu."
BASE COUNT      101 a    118 c    118 g    109 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BF562471/rev   ..

 Align seg 1/1  to reverse of: BF562471  from: 1  to: 446

     124 TrpGlyGlnIleValProThrLeu 131
         ||||||||||||||||||||||||
     429 TGGGGGCAAATTGTGCCAACACTG 406

seq_name: gb_gss:AQ600079

seq_documentation_block:
LOCUS       AQ600079                446 bp    DNA     linear   GSS 10-JUN-1999
DEFINITION  HS_5354_B2_G05_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=930 Col=10 Row=N, DNA sequence.
ACCESSION   AQ600079
VERSION     AQ600079.1  GI:5059996
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
```

```
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 930  row: N  column: 10
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 446.
FEATURES                Location/Qualifiers
     source             1. .446
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Plate=930 Col=10 Row=N"
                        /clone_lib="RPCI-11 Human Male BAC Library"
                        /sex="male"
                        /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                        Male blood DNA was isolated from one randomly chosen donor
                        and partially digested with a combination of EcoRI and
                        EcoRI Methylase. Size selected DNA was cloned into the
                        pBACe3.6 vector at EcoRI sites"
BASE COUNT      137 a     89 c    101 g    119 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AQ600079   ..

 Align seg 1/1  to: AQ600079  from: 1  to: 446

     239 SerIleCysArgHisValLeuPro 246
         |||||||||||||||||||||||||
     183 TCTATTTGCCGGCACGTGCTACCC 206

seq_name: gb_est1:AW215765

seq_documentation_block:
LOCUS       AW215765               455 bp    mRNA    linear   EST 06-DEC-1999
DEFINITION  up09g09.y1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:2651584 5',
            mRNA sequence.
ACCESSION   AW215765
VERSION     AW215765.1  GI:6526460
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            MGI:1032036
            Seq primer: -40RP from Gibco
            High quality sequence stop: 450.
FEATURES                Location/Qualifiers
     source             1. .455
                        /organism="Mus musculus"
                        /strain="CZECH II"
                        /db_xref="taxon:10090"
                        /clone="IMAGE:2651584"
                        /clone_lib="NCI_CGAP_Lu30"
                        /tissue_type="tumor, metastatic to mammary"
                        /lab_host="DH10B"
                        /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                        Site_2: SalI; transgenic model WNT-1, expression driven by
                        MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                        dT.   Library constructed by Life Technologies.
                        Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      77 a    144 c    138 g     96 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AW215765/rev   ..

 Align seg 1/1  to reverse of: AW215765  from: 1  to: 455

     139 SerArgGlyIleLeuProAlaAsn 146
         ||||||||||||||||||||||||
     383 TCACGAGGGATCTTGCCAGCCAAC 360

seq_name: gb_est1:AA509926

seq_documentation_block:
LOCUS       AA509926               456 bp    mRNA    linear   EST 08-JUL-1997
DEFINITION  vg30c10.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:862866 5', mRNA sequence.
ACCESSION   AA509926
VERSION     AA509926.1  GI:2247780
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:506954
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 425.
FEATURES                Location/Qualifiers
     source             1. .456
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="IMAGE:862866"
                        /clone_lib="Soares_mammary_gland_NbMMG"
                        /sex="male"
                        /tissue_type="mammary gland"
                        /dev_stage="4 weeks"
                        /lab_host="DH10B"
                        /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                        ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                        RI; 1st strand cDNA was primed with a Not I - oligo(dT)
```

primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 119 a 131 c 140 g 66 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AA509926/rev ..

Align seg 1/1 to reverse of: AA509926 from: 1 to: 456

101 LeuLeuThrThrAlaArgSerTrp 108
||||||||||||||||||||||||
187 CTTCTAACAACTGCGCGTTCATGG 164

seq_name: gb_gss:B45889

seq_documentation_block:
LOCUS B45889 458 bp DNA linear GSS 21-OCT-1997
DEFINITION HS-1062-B1-C12-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 784 Col=23 Row=F, DNA sequence.
ACCESSION B45889
VERSION B45889.1 GI:2550723
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 458)
AUTHORS Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
JOURNAL Unpublished (1997)
COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackron@u.washington.edu
Sequence Tagged Connector
Plate: CT 784 row: F column: 23
Class: BAC ends
High quality sequence stop: 458.
FEATURES Location/Qualifiers
source 1. .458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 784 Col=23 Row=F"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 128 a 102 c 85 g 143 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x B45889/rev ..

Align seg 1/1 to reverse of: B45889 from: 1 to: 458

8 PheIleThrThrLeuIleSerSer 15
|||||||||||||||||||||||||
347 TTTATCACAACATTGATTTCCTCC 324

seq_name: gb_est1:AA874458

seq_documentation_block:
LOCUS AA874458 459 bp mRNA linear EST 19-MAR-1998
DEFINITION vx80e06.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:1281538 5' similar to TR:Q61882 Q61882 SERINE/THREONINE KINASE 5 ;, mRNA sequence.
ACCESSION AA874458
VERSION AA874458.1 GI:2979147
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 459)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:673338
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 441.
FEATURES Location/Qualifiers
source 1. .459
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1281538"
/clone_lib="Soares_thymus_2NbMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 91 a 128 c 129 g 111 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
 US-09-674-779-2 x AA874458/rev   ..

 Align seg 1/1  to reverse of: AA874458  from: 1  to: 459

     122 HisLeuTrpGlyGlnIleValPro 129
         |||||||||||||||||||||||||
     138 CATCTCTGGGGGCAGATAGTACCA 115

seq_name: gb_est2:BI514220

seq_documentation_block:
LOCUS       BI514220                 464 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160014B10H08 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160014B10H08 5', mRNA sequence.
ACCESSION   BI514220
VERSION     BI514220.1  GI:15364594
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 464   Std Error: 0.00
            Plate: BB160014B10  row: H  column: 08
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Apis mellifera"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB160014B10H08"
                     /clone_lib="Bee Brain Normalized Library, BB16"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; The BB16 library was contributed by the
                     Soares laboratory and it was  constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G.  and Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared from dissected brains of adult worker bees of
                     various ages  and various behavioral groups. "
BASE COUNT      155 a     73 c     99 g    137 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI514220   ..

 Align seg 1/1  to: BI514220  from: 1  to: 464

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||||
      62 CAAAATTTTGGTTTGGGATTATAT 85

seq_name: gb_est2:BF565248

seq_documentation_block:
LOCUS       BF565248                 465 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  UI-R-BO1-ajj-c-10-0-UI.r1 UI-R-BO1 Rattus norvegicus cDNA clone
            UI-R-BO1-ajj-c-10-0-UI 5', mRNA sequence.
ACCESSION   BF565248
VERSION     BF565248.1  GI:11674978
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1798296
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BO1-ajj-c-10-0-UI"
                     /clone_lib="UI-R-BO1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The library
                     (UI-R-BO1) is a subtracted library derived from a mixture
                     of the following tissues: thalamus, cerebellum,
                     hypothalamus, medulla, pons, midbrain, cerebral cortex,
                     corpus striatum and hippocampus. For a detailed
                     description of the library from which this clone was
                     derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
BASE COUNT      109 a    148 c    147 g     60 t      1 others
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-674-779-2 x BF565248/rev   ..

 Align seg 1/1  to reverse of: BF565248  from: 1  to: 465

     101 LeuLeuThrThrAlaArgSerTrp 108
         ||||||||||||||||||||||||
     260 CTTCTAACAACTGCGCGTTCATGG 237

seq_name: gb_est2:BI513342

seq_documentation_block:
LOCUS       BI513342                477 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160012B10B07 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160012B10B07 5', mRNA sequence.
ACCESSION   BI513342
VERSION     BI513342.1  GI:15363716
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 477   Std Error: 0.00
            Plate: BB160012B10  row: B  column: 07
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Apis mellifera"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB160012B10B07"
                     /clone_lib="Bee Brain Normalized Library, BB16"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; The BB16 library was contributed by the
                     Soares laboratory and it was  constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G.  and Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared from dissected brains of adult worker bees of
                     various ages  and various behavioral groups. "
BASE COUNT      159 a     76 c    100 g    142 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI513342   ..

 Align seg 1/1  to: BI513342  from: 1  to: 477

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||||
      74 CAAAATTTTGGTTTGGGATTATAT 97

seq_name: gb_gss:BH105280

seq_documentation_block:
LOCUS       BH105280                487 bp    DNA     linear   GSS 19-JUL-2001
DEFINITION  RPCI-24-351K23.TJ RPCI-24 Mus musculus genomic clone RPCI-24-351K23
            , DNA sequence.
ACCESSION   BH105280
VERSION     BH105280.1  GI:14934069
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-351K23.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 351  row: K  column: 23
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-351K23"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      125 a    103 c    141 g    118 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BH105280/rev   ..

 Align seg 1/1  to reverse of: BH105280  from: 1  to: 487
```

```
     169 LysHisLeuThrAsnSerAlaIle 176
         ||||||||||||||||||||||||
     260 AAGCACCTTACTAACTCAGCCATC 237

seq_name: gb_est2:BI512001

seq_documentation_block:
LOCUS       BI512001                489 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160007B20D07 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160007B20D07 5', mRNA sequence.
ACCESSION   BI512001
VERSION     BI512001.1  GI:15362375
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 489   Std Error: 0.00
            Plate: BB160007B20  row: D  column: 07
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 489.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Apis mellifera"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB160007B20D07"
                     /clone_lib="Bee Brain Normalized Library, BB16"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; The BB16 library was contributed by the
                     Soares laboratory and it was  constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G.  and Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared from dissected brains of adult worker bees of
                     various ages  and various behavioral groups. "
BASE COUNT      164 a     79 c    105 g    141 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI512001   ..
```

```
 Align seg 1/1  to: BI512001  from: 1  to: 489

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||||
      67 CAAAATTTTGGTTTGGGATTATAT 90

seq_name: gb_est2:BI515132

seq_documentation_block:
LOCUS       BI515132                516 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160017B10C08 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160017B10C08 5', mRNA sequence.
ACCESSION   BI515132
VERSION     BI515132.1  GI:15365506
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 516   Std Error: 0.00
            Plate: BB160017B10  row: C  column: 08
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Apis mellifera"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB160017B10C08"
                     /clone_lib="Bee Brain Normalized Library, BB16"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; The BB16 library was contributed by the
                     Soares laboratory and it was  constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G.  and Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared from dissected brains of adult worker bees of
                     various ages  and various behavioral groups. "
BASE COUNT      152 a    100 c    122 g    142 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-674-779-2 x BI515132   ..

 Align seg 1/1  to: BI515132  from: 1  to: 516

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||||
      75 CAGAATTTCGGATTAGGCCTTTAT 98

seq_name: gb_gss:B48301

seq_documentation_block:
LOCUS       B48301                 528 bp    DNA     linear   GSS 08-APR-1999
DEFINITION  RPCI11-6P4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-6P4, DNA
            sequence.
ACCESSION   B48301
VERSION     B48301.1  GI:2600538
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="GDB:7502283"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-6P4"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      153 a    114 c     76 g    185 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x B48301   ..

 Align seg 1/1  to: B48301  from: 1  to: 528

     240 IleCysArgHisValLeuProLys 247
         |||||||||||||||||||||||
      11 ATCTGTAGACATGTATTACCAAAA 34

seq_name: gb_est2:BI643242
```

```
seq_documentation_block:
LOCUS       BI643242               533 bp    mRNA    linear   EST 10-SEP-2001
DEFINITION  RS4_C01 Sugar beet root cDNA library (subtracted) Beta vulgaris
            cDNA 5', mRNA sequence.
ACCESSION   BI643242
VERSION     BI643242.1  GI:15545452
KEYWORDS    EST.
SOURCE      Beta vulgaris.
  ORGANISM  Beta vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   de los Reyes,B.G., McGrath,J.M. and Myers,S.
  TITLE     Differential gene expression in sugar beet (Beta vulgaris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: J. Mitchell McGrath
            Sugar Beet Genetics Lab., USDA/ARS/Sugar Beet and Bean Research
            Unit
            Michigan State University
            494 Plant and Soil Sci. Bldg., East Lansing MI 48824, USA
            Tel: (517)-432-2355
            Fax: (517)-337-6782
            Email: mitchmcg@pilot.msu.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .533
                     /organism="Beta vulgaris"
                     /cultivar="USH20"
                     /db_xref="taxon:161934"
                     /clone_lib="Sugar beet root cDNA library (subtracted)"
                     /tissue_type="mature root"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBK-CMV; Site_1: EcoRI;
                     Site_2: XhoI; cDNAs were derived from reverse
                     transcription of mRNA samples from mature root tissue. The
                     mRNA pool was used as template for double stranded cDNA
                     synthesis using the Stratagene pBluescript XR cDNA
                     synthesis and library kit. The resulting cDNA was used as
                     a tester for subtraction against a driver cDNA population
                     derived from leaf, inflorescence, and 4-day old germinated
                     seedlings. Subtraction was performed using the Invitrogen
                     Photobiotin-Streptavidin subtractor kit. The cDNA library
                     was generated by directional ligation of the subtracted
                     cDNAs in the EcoRI and XhoI sites of pBK-CMV vector
                     (Stratagene). Putative ID reported when E scores were
                     better than e-10 via nr database and blastx."
BASE COUNT      193 a     79 c    130 g    131 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI643242/rev   ..

 Align seg 1/1  to reverse of: BI643242  from: 1  to: 533

      89 SerArgLeuGlyAsnTyrLeuPro 96
         |||||||||||||||||||||||||
     174 AGTCGCTTAGGAAACTACTTGCCA 151

seq_name: gb_est2:BI517071

seq_documentation_block:
LOCUS       BI517071               535 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160024A20E01 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160024A20E01 5', mRNA sequence.
ACCESSION   BI517071
```

```
VERSION     BI517071.1  GI:15367445
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 535    Std Error: 0.00
            Plate: BB160024A20  row: E  column: 01
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 535.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Apis mellifera"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB160024A20E01"
                     /clone_lib="Bee Brain Normalized Library, BB16"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; The BB16 library was contributed by the
                     Soares laboratory and it was  constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G.  and Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared from dissected brains of adult worker bees of
                     various ages  and various behavioral groups. "
BASE COUNT      183 a    103 c    121 g    128 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI517071   ..

 Align seg 1/1  to: BI517071  from: 1  to: 535

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||||
      17 CAGAATTTCGGATTAGGCCTTTAT 40

seq_name: gb_est1:AI995165

seq_documentation_block:
LOCUS       AI995165                 544 bp    mRNA    linear   EST 08-SEP-1999
DEFINITION  701502426 A. thaliana, Ohio State clone set Arabidopsis thaliana
            cDNA clone 701502426, mRNA sequence.
ACCESSION   AI995165
VERSION     AI995165.1  GI:5842070
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
            Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
            Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L.,
            Turner,C., Krikorian,S., Elder,L. and Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="701502426"
                     /clone_lib="A. thaliana, Ohio State clone set"
                     /note="cDNA library was made from selected clones from the
                     Arabidopsis thaliana Ohio  State clone set."
BASE COUNT      165 a     98 c    141 g    140 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AI995165/rev   ..

 Align seg 1/1  to reverse of: AI995165  from: 1  to: 544

     144 ProAlaAsnThrGlnIleArgSer 151
         |||||||||||||||||||||||
     483 CCGGCCAATACACAGATCCGATCC 460

seq_name: gb_gss:AQ266185

seq_documentation_block:
LOCUS       AQ266185                 579 bp    DNA     linear   GSS 27-OCT-1998
DEFINITION  CITBI-E1-2502L22.TF CITBI-E1 Homo sapiens genomic clone 2502L22,
            DNA sequence.
ACCESSION   AQ266185
VERSION     AQ266185.1  GI:3790779
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
```

```
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .579
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="2502L22"
                      /clone_lib="CITBI-E1"
                      /sex="male"
                      /cell_type="sperm"
                      /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                      CalTech Human BAC Library D"
BASE COUNT       160 a    114 c    111 g    194 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AQ266185    ..

 Align seg 1/1  to: AQ266185  from: 1  to: 579

       8 PheIleThrThrLeuIleSerSer 15
         |||||||||||||||||||||||
     379 TTCATAACCACTCTTATTTCATCT 402

seq_name: gb_est2:BI515482

seq_documentation_block:
LOCUS       BI515482                 580 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  BB160019B10B04 Bee Brain Normalized Library, BB16 Apis mellifera
            cDNA clone BB160019B10B04 5', mRNA sequence.
ACCESSION   BI515482
VERSION     BI515482.1  GI:15365856
KEYWORDS    EST.
SOURCE      honeybee.
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
            ; Apoidea; Apidae; Apis.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
            Smoller,D. and Robinson,G.E.
  TITLE     An Expressed Sequence Tag Resource for Studies of Brain and
            Behavior in the Honey Bee
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 580   Std Error: 0.00
            Plate: BB160019B10  row: B  column: 04
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 580.
FEATURES              Location/Qualifiers
     source           1. .580
                      /organism="Apis mellifera"
                      /strain="mixed strains of European bees, predominantly
                      A.m. ligustica"
                      /db_xref="taxon:7460"
                      /clone="BB160019B10B04"
                      /clone_lib="Bee Brain Normalized Library, BB16"
                      /sex="female"
                      /tissue_type="brain"
                      /dev_stage="adult worker honey bee"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                      Site_2: Not1; The BB16 library was contributed by the
                      Soares laboratory and it was  constructed and normalized
                      as described by Bonaldo, M.F., Lennon, G.  and Soares,
                      M.B. (1996), Genome Research 6(9): 791-806. RNA was
                      prepared from dissected brains of adult worker bees of
                      various ages  and various behavioral groups. "
BASE COUNT       189 a    113 c    134 g    144 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BI515482    ..

 Align seg 1/1  to: BI515482  from: 1  to: 580

     208 GlnAsnPheGlyLeuGlyLeuTyr 215
         |||||||||||||||||||||||||
      62 CAGAATTTCGGATTAGGCCTTTAT 85
```

ON of: US-09-674-77[redacted]o: GenEmbl:* out_format : pfs

Date: Sep 16, 2002 6:52 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124121_25832/app_query.fasta_1.310
-DB=GenEmbl -QFMT=fastap -SUFFIX=std.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_4034 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 1868.280000

score_list:

| Sequence | Strd | Orig | ZScore | EScore | Len | ! Documentation .. |
|---|---|---|---|---|---|---|
| gb_pat:AX081145 | + | 1355.00 | 2163.56 | 3.8e-112 | 753 | ! AX081145 Sequence 1 from Patent |
| gb_pat:AX067442 | + | 1355.00 | 2131.81 | 2.2e-110 | 23210 | ! AX067442 Sequence 17 from Pat |
| gb_ba:AF400582 | + | 375.00 | 571.09 | 1.9e-23 | 8367 | ! AF400582 Acinetobacter sp. ADF |
| gb_ba:SME591791 | - | 127.00 | 139.41 | 21.10 | 340900 | ! AL591791 Sinorhizobium melil |
| gb_ba:AP003009 | + | 112.00 | 115.41 | 458.27 | 339681 | ! AP003009 Mesorhizobium loti |
| gb_htg:AC046147 | - | 110.00 | 121.81 | 201.66 | 120405 | ! AC046147 Mus musculus chromo |
| gb_htg:AC073821 | - | 110.00 | 120.07 | 252.19 | 145347 | ! AC073821 Mus musculus clone |
| gb_htg:AC073767 | + | 110.00 | 117.05 | 371.47 | 201377 | ! AC073767 Mus musculus clone |
| gb_ba:AP003000 | - | 106.50 | 106.37 | 1.5e+03 | 348077 | ! AP003000 Mesorhizobium loti |
| gb_pr:AC068614 | - | 106.00 | 112.50 | 665.59 | 164706 | ! AC068614 Homo sapiens BAC cl |
| gb_htg:AC098619 | + | 105.50 | 110.67 | 842.27 | 184172 | ! AC098619 Rattus norvegicus c |
| gb_htg:AC020643 | - | 103.50 | 108.62 | 1.1e+03 | 162527 | ! AC020643 Homo sapiens chromo |
| gb_htg:AC027289 | - | 103.50 | 108.00 | 1.2e+03 | 173797 | ! AC027289 Homo sapiens chromo |
| gb_ba:AE005826 | + | 103.00 | 131.55 | 57.82 | 12528 | ! AE005826 Caulobacter crescent |
| gb_htg:AC003656_4 | - | 102.00 | 109.83 | 937.46 | 110000 | ! Continuation (5 of 8) of AC |
| gb_pr:AP001053 | - | 102.00 | 106.25 | 1.5e+03 | 161920 | ! AP001053 Homo sapiens genomi |
| gb_htg:AC079571 | - | 102.00 | 102.41 | 2.4e+03 | 245130 | ! AC079571 Mus musculus clone |
| gb_pr:AP001752 | - | 102.00 | 99.38 | 3.6e+03 | 340000 | ! AP001752 Homo sapiens genomi |
| gb_pr:AC078925 | + | 101.50 | 105.96 | 1.5e+03 | 153284 | ! AC078925 Homo sapiens 12q BA |
| gb_pr:AL133414 | - | 101.50 | 105.61 | 1.6e+03 | 159169 | ! AL133414 Human DNA sequence |
| gb_pr:AC069304 | - | 101.50 | 105.49 | 1.6e+03 | 161247 | ! AC069304 Homo sapiens BAC cl |
| gb_ba:AE008307 | + | 101.00 | 128.89 | 81.34 | 11814 | ! AE008307 Agrobacterium tumefa |
| gb_ba:AE009307 | - | 101.00 | 128.81 | 82.15 | 11913 | ! AE009307 Agrobacterium tumefa |
| gb_pr:AC006925 | + | 100.50 | 103.06 | 2.2e+03 | 176186 | ! AC006925 Homo sapiens chromo |
| gb_htg:AC106530 | - | 100.50 | 102.15 | 2.5e+03 | 194508 | ! AC106530 Rattus norvegicus c |
| gb_in:CAE223616 | + | 100.00 | 136.47 | 30.77 | 4383 | ! AJ223616 Caehorhabditis elegan |
| gb_pr:HS833B7 | - | 100.00 | 108.84 | 1.1e+03 | 86574 | ! AL008637 Human DNA sequence f |
| gb_pr:AC083964 | + | 100.00 | 103.09 | 2.2e+03 | 161160 | ! AC083964 Homo sapiens chromo |
| gb_htg:AC090135 | - | 100.00 | 101.56 | 2.7e+03 | 190173 | ! AC090135 Homo sapiens chromo |
| gb_ba:AP003008 | - | 100.00 | 95.99 | 5.5e+03 | 346820 | ! AP003008 Mesorhizobium loti |
| gb_htg:AC105039 | - | 99.50 | 103.22 | 2.2e+03 | 145667 | ! AC105039 Homo sapiens chromo |
| gb_htg:AC016944 | - | 99.50 | 102.42 | 2.4e+03 | 158886 | ! AC016944 Homo sapiens chromo |
| gb_pr:AL139038 | - | 99.00 | 102.74 | 2.3e+03 | 140756 | ! AL139038 Human DNA sequence |
| gb_htg:AC073525 | - | 99.00 | 99.67 | 3.4e+03 | 196048 | ! AC073525 Homo sapiens chromo |
| gb_htg:AC103346 | + | 98.50 | 116.24 | 412.26 | 30062 | ! AC103346 Rattus norvegicus cl |
| gb_htg:AC013934 | - | 98.50 | 100.22 | 3.2e+03 | 169516 | ! AC013934 Drosophila melanoga |
| gb_in:AC009905 | + | 98.50 | 100.16 | 3.2e+03 | 170662 | ! AC009905 Drosophila melanoga |
| gb_in:AC010121 | + | 98.50 | 100.05 | 3.3e+03 | 172529 | ! AC010121 Drosophila melanoga |
| gb_htg:AC087566 | + | 98.50 | 97.32 | 4.7e+03 | 231912 | ! AC087566 Mus musculus clone |
| gb_in:AE003655 | + | 98.50 | 96.48 | 5.2e+03 | 253753 | ! AE003655 Drosophila melanoga |
| gb_pr:HS278N12 | + | 98.00 | 106.54 | 1.4e+03 | 78533 | ! AL033376 Human DNA sequenc |
| gb_htg:AC097543 | + | 98.00 | 100.78 | 3.0e+03 | 146321 | ! AC097543 Rattus norvegicu |
| gb_in:PMHISH2A | + | 97.50 | 148.33 | 6.72 | 790 | ! X01346 Psammechinus miliaris |
| gb_in:SUPHISB4 | + | 97.50 | 147.42 | 7.56 | 872 | ! M10559 P.miliaris histone H2 |
| em_htgo_inv:AC034780 | + | 97.50 | 145.56 | 9.59 | 1066 | ! Ac034780 Giardia intestinal |

seq_name: gb_pat:AX081145

seq_documentation_block:
LOCUS AX081145 753 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 1 from Patent WO0109335.
ACCESSION AX081145
VERSION AX081145.1 GI:13170040
KEYWORDS .
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 753)
AUTHORS Thonnard,J.
TITLE Moraxella catarrhalis antigen basb120
JOURNAL Patent: WO 0109335-A 1 08-FEB-2001;
SmithKline Beecham Biologicals s.a. (BE)
FEATURES Location/Qualifiers
source 1. .753
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 239 a 166 c 152 g 196 t
ORIGIN

alignment_scores:
Quality: 1355.00 Length: 250
Ratio: 5.420 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AX081145 ..

Align seg 1/1 to: AX081145 from: 1 to: 753

```
      1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 50

     17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     51 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 100

     34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    101 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 150

     51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    151 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 200

     67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 250

     84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    251 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 300

    101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    301 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 350

    117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    351 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 400
```

```
    134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 450

    151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 500

    167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 550

    184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    551 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 600

    201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    601 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 650

    217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    651 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 700

    234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    701 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 750
```

seq_name: gb_pat:AX067442

```
seq_documentation_block:
LOCUS       AX067442              23210 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 17 from Patent WO0078968.
ACCESSION   AX067442
VERSION     AX067442.1  GI:12545062
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 23210)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 17 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .23210
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
     unsure          28
                     /note="a or g or c or t, unknown, or other"
BASE COUNT     7108 a    4482 c    4951 g    6668 t       1 others
ORIGIN
```

```
alignment_scores:
            Quality: 1355.00              Length:    250
              Ratio:  5.420                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AX067442   ..

 Align seg 1/1  to: AX067442  from: 1  to: 23210

      1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11483 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 11532

     17 uValAlaCysSe[redacted]uProIleProThrAsnProGlnValSerProIleL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11533 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 11582
```

```
     34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11583 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 11632

     51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11633 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 11682

     67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11683 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 11732

     84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11733 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 11782

    101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11783 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 11832

    117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11833 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 11882

    134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11883 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 11932

    151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11933 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 11982

    167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11983 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 12032

    184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12033 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 12082

    201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12083 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 12132

    217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12133 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 12182

    234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12183 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 12232
```

seq_name: gb_ba:AF400582

```
seq_documentation_block:
LOCUS       AF400582              8367 bp    DNA     linear   BCT 08-NOV-2001
DEFINITION  Acinetobacter sp. ADP1 mismatch repair protein (mutS), 7-Fe
            ferredoxin (fdxA), and O-methyltransferase-like protein genes,
            complete cds; and unknown genes.
ACCESSION   AF400582
VERSION     AF400582.1  GI:15217080
KEYWORDS    .
SOURCE      Acinetobacter sp. ADP1.
  ORGANISM  Acinetobacter sp. ADP1
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Acinetobacter.
REFERENCE   1  (bases 1 to 8367)
  AUTHORS   Young,D.M. and Ornston,L.N.
  TITLE     Functions of the Mismatch Repair Gene mutS from Acinetobacter sp.
            Strain ADP1
  JOURNAL   J. Bacteriol. 183 (23), 6822-6831 (2001)
  MEDLINE   21555098
   PUBMED   11698371
```

```
REFERENCE   2  (bases 1 to 8367)
  AUTHORS   Young,D.M. and Ornston,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2001) Molecular, Cellular, and Developmental
            Biology, Yale University, P.O. Box 208103, New Haven, CT
            06520-8103, USA
FEATURES             Location/Qualifiers
     source          1. .8367
                     /organism="Acinetobacter sp. ADP1"
                     /strain="ADP1; BD413"
                     /db_xref="taxon:62977"
     CDS             complement(12. .725)
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92492.1"
                     /db_xref="GI:15217081"
                     /translation="MTTLIQTPLSNAVNNFIDLNREKAEFAFEVFRETKTITANGTVN
                     FVERGPDNTLITFNYAGPWSKDKKSKVLVETLDGEVLLGRKTAGGRYNKLFREHVDVT
                     TISHVHAPNLGAWAQTHRTLPIRYVPVQRFHLFKELPVYIDRRQAEVDFILDQLKVDI
                     EYRAILEANGGATVWGRQGLQETAEFILLLEEGAQIQLLAELIGGSREYGQGVLHQQW
                     KMSGLTEKAKSLGLLPLNN"
     CDS             complement(739. .1797)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92493.1"
                     /db_xref="GI:15217082"
                     /translation="MSLVVNKKENSKEFVKNILSEIWVTRCPVPTATGISLKRGTLVK
                     KFEKHGIDVAVLQDARKGLSVHHFDHQLPALFREGGNVPALAARSEGAPSRLIGLTWI
                     DEWQVIIVRPDSDISQPEHLLNAKVALPEYADKRAGSIFRAMSLQGIRGALQLANLSF
                     RDVNFIEVPERAGRKENRDASQYWSGLDALITGQVDAVYVKGASAVEAATKRGLKVGI
                     DLDQFPDITSRVNNGTPRPITVHEDLLEDHFDVVVDFLEELFDTADWAQDHLDEVLTI
                     LQDETIAGAESVATAYRNNFHKSLHPSLSTERLELLKQQKDFLWLHGFLERDVNIDNW
                     VDHRPLLEVLKRRQNKQL"
     gene            2592. .5237
                     /gene="mutS"
     CDS             2592. .5237
                     /gene="mutS"
                     /note="MutS"
                     /codon_start=1
                     /transl_table=11
                     /product="mismatch repair protein"
                     /protein_id="AAK92494.1"
                     /db_xref="GI:15217083"
                     /translation="MNSTETMADLSSYTPMMQQYFKVKLEHCHALLFYRMGDFYELFF
                     DDARKAAKFLGITLTHRGKANGEPIPMAGVPYHAAEGYLARLVRAGQTVAICEQVGEG
                     ENAGSRCKAPMERKVVRIITPGTITDDALLGSYQSSNLVALCIQQNKIGIALLDLSAS
                     IFKVQQHEFKTEQLYIELARLMPSEIVVDEDLVDQNILEQIKKQIECSITKRPNVDFN
                     LNNAQKTLCDQFGVTTLSGFGIDHLPLAKAAAAALIHYAKETQKTALPHIRSIQLEQS
                     TDFIALDPITRRNLEIIDPLFEHGTSLFNLINDCQTAMGGRLLSRILMQPIRDTAILD
                     ARLDATEQLLIGYHESPVRLVLKEIGDIERVLSRVALGTARPRDLVQLRQACAQIPFL
                     RHALLPAIQTKSSKLLNELDHELGDFKDLHQLLLSAIVENPPVLLRDGNVIAEGYDAE
                     LDELRKIRDHAGQFLVDLEIKERESSGIPTLKIGYNRVSGYYIELTRAQAEQAPAYYI
                     RRQTLKNAERYITPELKSFEDKVLSSESRALAREKLLFESLLETLKKNIANLQMMSSA
                     IAQIDVLANFAHQARLQNWNRPDFSPEIGIKIHDGRHPVVEALSKTPYTPNDTFLDSQ
                     HRMAIITGPNMGGKSTFMRQTALISLLAYCGSYVPAKSAKLGPIDRIFTRIGSADDLS
                     TGKSTFMVEMTETSQILHHATHQSLVLMDEVGRGTSTYDGLSLAWACVLDLTKRIKCL
                     CLFATHYFELTELGSESAIDNYHVTAKEMNGNLILLHKVQHGPASQSHGLQVAKLAGI
                     PASVIKEAQKRLKILEKQQQQYFQTAVQNDLFAIPELPQIDETKIEIKKTSPVLDHLK
                     QLDVDSLTPRQALEALYQLKDQLEH"
     CDS             5493. .5942
                     /note="ORF1; similar to FAD-binding domain of AppA redox
                     regulator from Rhodobacter sphaerodies"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92495.1"
                     /db_xref="GI:15217084"
                     /translation="MAYVRLLYVSKTAKQHHEIKHDLMEILTTAIRFNSINKIKGVLY
                     YGNGYFVQCLEGEEKVIYDVFYNRIIKDSRHQNCEILSCKAVNELLFQKWSMKFAPLN
                     KKLRDFFQQYHIDEFNPYLLLTHQTTEPFIELLSQEPHQEVDPYPHSS"
     gene            6267. .6590
                     /gene="fdxA"
     CDS             6267. .6590
                     /gene="fdxA"
                     /note="FdxA; similar to AvFdI from Azotobacter vinelandii"
                     /codon_start=1
                     /transl_table=11
                     /product="7-Fe ferredoxin"
                     /protein_id="AAK92496.1"
                     /db_xref="GI:15217085"
                     /translation="MTFVVTENCIKCKYQDCVEVCPVDCFYEGPNFLVINPDECIDCA
                     LCEPECPANAIFSEDELPEGQEVFIELNADLSQKWPNITQIGDQPADREEWNGKPDKL
                     QYLEK"
     CDS             6717. .7439
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK92497.1"
                     /db_xref="GI:15217086"
                     /translation="MSQFLKGVLAISIIPIMMVGCTTAPQKKGKILDLPKGHRIHIPQ
                     ETVTIERPVRPKVTPNSYLNWLQNPSNRIQIQEYEQFLSRNDVANIVPNYELLKTARD
                     WQKCGRSEYAVPNRELWNNAIPTLRVLKYLVAAKILTDFEVTSVYRDLPLNQCAGGAN
                     SSRHLFNSAIDFRIGPEIPQPQDYAFIENTKFKLCQFWNQHGQSLNMGLGLYSSGQIH
                     IDTQGYRTWGPDHSRNSSMCNY"
     CDS             7593. .8261
                     /note="ORF3; member of O-methyltransferase family"
                     /codon_start=1
                     /transl_table=11
                     /product="O-methyltransferase-like protein"
                     /protein_id="AAK92498.1"
                     /db_xref="GI:15217087"
                     /translation="MQQLWTDIDHYIESNLIPEDSNLKFALQNTEDQGFPDHLAVAPN
                     QGMLLQMLIQMNQCKRVLELGTFAAYSTIWLARALPKDGYLMTIEGRDTHAALAQQNI
                     DHANLPVNIELKVGRAADILSKIDTDTFEPFDFIFIDADKQSYPEYLDLSLKLSHSGT
                     VIVLDNVIRAGALIDLDNQRPSIEGIRLTFEALQNHPQLLSCTALQTVGSKGHDGFAI
                     AIVK"
BASE COUNT     2656 a   1606 c   1445 g   2660 t
ORIGIN


alignment_scores:
            Quality:  375.00            Length:    248
              Ratio:  2.358               Gaps:     10
 Percent Similarity:  64.113   Percent Identity: 35.887

alignment_block:
 US-09-674-779-2 x AF400582   ..

 Align seg 1/1  to: AF400582  from: 1  to: 8367

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ::::::|||      |||:::|||:::|||:::            |||
    6750 AGTATCATCCCGATAATGATGGTTGGCTGTACC.............ACTGC 6787

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
         ||||||                                    |||   |
    6788 TCCTCAA.................................AAAAAAGGAA 6804

      44 ysIleGlyAsp.............HisHisThrHisGluHisAspGluSer 56
         |||||   |||            |||:::   |||      :::|||:::
    6805 AAATCTTAGACTTACCAAAGGGACATCGTATTCATATTCCTCAGGAAACT 6854

      57 ValSer.....................HisValGlyLeuGlnAlaHisPh 66
         |||:::                        |||      ::::::::::
    6855 GTGACTATAGAACGTCCTGTTCGCCCTAAAGTAACACCAAACTCTTATTT 6904

      66 eGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgT 83
         :   :::|||||||||   :::   :::::::::   ::::::   :::|
    6905 G...AATTGGCTGCAA...AATCCGTCTAATCGGATACAAATACAAGAAT 6948

      83 yrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSer 99
```

```
      ||:::   :::|||:::   :::            :::   |||
   6949 ATGAGCAATTTTTATCACGAAACGATGTAGCAAATATTGTTCCAAATTAC 6998

    100 GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTy 116
        :::||||||   ||||||||||:::||||||   ||||||::::::   ||
   6999 GAATTATTGAAAACAGCTCGCGATTGGCAAAAATGTGGACGATCCGAATA 7048

    116 rGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisL 133
        |   :::|||   ::::::||||||::::::   :::||||||||||:::
   7049 TGCGGTTCCCAATAGAGAGCTTTGGAATAATGCCATTCCTACATTACGTG 7098

    133 euTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIle 149
        ::   :::   |||   :::      ||||||   ::::::   ::::::
   7099 TGCTCAAATATTTAGTGGCAGCAAAAATTTTG...ACTGATTTTGAGGTG 7145

    150 ArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAl 166
        ||||||||||||:::      ||||||||||||||||||||||||||
   7146 ACATCTGTTTATCGTGATTTACCATTAAATCAGTGTGCTGGTGGCGCGAA 7195

    166 aMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpVal...ProA 182
        |||:::||||||   ||||||||||||||:::   :::   |||:
   7196 TTCTTCCCGGCATTTGTTTAATTCGGCTATCGATTTTCGTATTGGTCCCG 7245

    182 spLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsn..... 195
        :::::      :::   |||           :::::::|||
   7246 AAATTCCT...CAACCGCAAGATTATGCTTTTATTGAAAATACCAAATTT 7292

    196 ArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLe 212
        :::||||||||||:::|||   :::||||||::::::   |||:::|||||
   7293 AAGCTTTGCCAATTTTGGAATCAACATGGACAAAGTCTGAATATGGGACT 7342

    212 uGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgL 229
        ||||||||||::::::|||   ||||||:::|||||||||||||::|||
   7343 AGGTCTGTACAGTTCAGGTCAAATTCATATTGACACTCAAGGCTATCGTA 7392

    229 ysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
        ||||||   :::   |||:::::::::|||:::|||::::::
   7393 CATGGGGGCCAGACCATAGCCGAAACAGTTCAATGTGTAATTAC 7436
```

seq_name: gb_ba:SME591791

seq_documentation_block:

LOCUS SME591791 340900 bp DNA linear BCT 16-AUG-2001
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 10/12.
ACCESSION AL591791 AL591688
VERSION AL591791.1 GI:15075538
KEYWORDS .
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 340900)
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J., Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S., Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D., Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U., Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
TITLE From the Cover: Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021
JOURNAL Proceedings of the National Academy of Sciences of the United States of America. 98 (17), 9877-9882 (2001)
PUBMED 11481430
REFERENCE 2 (bases 1 to 340900)
AUTHORS Gouzy,J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO EU Consortium
COMMENT MELILO EU Consortium:
Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet, France, Laboratoire de Genetique et Developpement UMR6061-CNRS, Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes, France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany, Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany, Unite de Biochimie physiologique, Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6, B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr http://sequence.toulouse.inra.fr/meliloti.html.

FEATURES Location/Qualifiers

source 1. .340900
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"

gene 132. .518
/gene="SMc01981"

CDS 132. .518
/gene="SMc01981"
/function="small molecule metabolism; energy transfer; electron transport"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE CYTOCHROME C TRANSMEMBRANE PROTEIN"
/protein_id="CAC47094.1"
/db_xref="GI:15075539"
/translation="MRIALAALSFSALSLCPVAGLAQEGDAEAGATVFKKCATCHVID
KDQNKVGPSLQGVIGRTAGTHADFKYSQAMIDAGKGGLVWDDATLAEYLRNPRAKVKG
TKMVFPGLKKDEEIANVIAYLKQHPK"

gene 656. .1600
/gene="coxM OR SMc01982"

CDS 656. .1600
/gene="coxM OR SMc01982"
/EC_number="1.9.3.1"
/function="small molecule metabolism; energy transfer; electron transport"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE POLYPEPTIDE II TRANSMEMBRANE PROTEIN"
/protein_id="CAC47095.1"
/db_xref="GI:15075540"
/translation="MAVVVILVLLAVGSVLFHLLSPWWWTPIASNWNYIDNTITITFW
ITGIAFTAVVLFMAYCVLRFRHRPGNTAAYEPENRRLEGWLATGTTFGVAAMLAPGLF
VWNQFVTVPQDASEVEVIGQQWLWSFRLPGADGKLGTTETRDIAPENTLGVNRDDAAG
QDDIIIEGGELHLPVGKPVKMLLRSVDVLHDFYVPEFRAKMDMVPGMITYFWLTPTRT
GTFEILCAELCGVGHPQMRGTVVVDTEEDYQAWLAEQQTFSQLSASSETRAVPEKVCS
GFPSGIATEQGTGASALFKEERECFGPAATTVAASAAQ"

repeat_region complement(1457. .1513)
/note="Sm-5 OR SMc04646
REPEAT SM-5
predicted by Homology"
/evidence=not_experimental

gene 1634. .3415
/gene="coxN OR SMc01983"

CDS 1634. .3415
/gene="coxN OR SMc01983"
/EC_number="1.9.3.1"
/function="small molecule metabolism; energy transfer; electron transport"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology

```
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
                POLYPEPTIDE I TRANSMEMBRANE PROTEIN"
                /protein_id="CAC47096.1"
                /db_xref="GI:15075541"
                /translation="MMVDVRSGIGEALPPPEVEDVELYHPHSWWTRYVFSQDAKIIAI
                QYSMTAIAIGMVALVLSWLIRLQLGFPGTFELIDAERYYQFITMHGMIMVIYLLTALF
                LGGFGNYLIPLMVGARDMVFPYANMLSYWIYLLAVIVLAASFFTPGGPTGAGWTLYPP
                QAVLSGTPGGRDWGIIMMLSSLIIFVIGFTMGGLNYVVTVLQGRARGMTLMRLPLTVW
                GIFTATVMALLAFPALFVACVMMLFDRLLGTSFFMPAIVEMGEQLQYGGGSPILFQHL
                FWFFGHPEVYIVALPAFGIVSDLISTHARKNIFGYRMMVWAIVIIGGLSFIVWAHHMY
                VSGMNPYFGFFFATTTLIIAVPTAIKVYNWVLTLWRGNIHLTLPMLFALAFIVTFVNG
                GLTGLFLGNVVVDVPLSDTMFVVAHFHMVMGVAPIMVIFGAIYHWYPKITGRMLNEAM
                GQIHFWVTFIGAYAIFFPMHYLGLIGVPRRYHELGEASFVTTSIAELNAFISVMALLV
                GAAQIVFLFNLAWSLRHGREAGGNPWRATTLEWQTPETPPPHGNWGNELPVVYRWAYD
                YSVPGAPEDFIPQNQPTPGRLSHETVS"
gene            3412. .4110
                /gene="coxO OR SMc01984"
CDS             3412. .4110
                /gene="coxO OR SMc01984"
                /EC_number="1.9.3.-"
                /function="small molecule metabolism; energy transfer;
                electron transport"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE CYTOCHROME-C OXIDASE TRANSMEMBRANE
                PROTEIN"
                /protein_id="CAC47097.1"
                /db_xref="GI:15075542"
                /translation="MSIVAFFLAAIAAIIAWWLAGQRLTSRPWLEVGHFHDRRGATRL
                PPAKIGLGVFLAVVGALFSLAISAYFMRMASSDWGALPLPGLLWLNTGILAAGSITLH
                WTKVEAERRNDEAARIGLLAGLALGLAFLAGQLFAWRALSDAGYFLAGNPANSFFYLL
                TGMHGLHIIGGLFALGRVTAHASQTPLGNRTRLSIELCAIYWHFMLIVWLVLFALFAG
                WASGVIEFCRQLLT"
gene            4123. .4845
                /gene="coxP OR SMc01985"
CDS             4123. .4845
                /gene="coxP OR SMc01985"
                /EC_number="1.9.3.-"
                /function="small molecule metabolism; energy transfer;
                electron transport"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE CYTOCHROME-C OXIDASE PROTEIN"
                /protein_id="CAC47098.1"
                /db_xref="GI:15075543"
                /translation="MSAPMKKPAAETLPRPAGLAGLASDWASDQRTFKDVSWGKAMMW
                IFLLSDTFVFGCFLLAYMSARMSTSVPWPNPSEVFALEIGGTHMPLILIAIMTFVLIS
                SSGTMAMAVNYGYRRDRRKTAALMLLTALFGAAFVGMQAFEWSKLIAEGVRPWGNPWG
                AAQFGSTFFMITGFHGTHVTFGVIFLLIVARKVWRGDFETERRGFFTSRKGRYEIVEI
                TGLYWHFVDLVWVFIFAFFYLW"
gene            4857. .5216
                /gene="SMc01986"
CDS             4857. .5216
                /gene="SMc01986"
                /function="miscellaneous; hypothetical/partial homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                /protein_id="CAC47099.1"
                /db_xref="GI:15075544"
                /translation="MAHAETHPAQHAAAHTEHQQHAIKLYLLVWGLLFVLSALSYLVD
                YFGLQGYLRWSLILIFMMLKAGLIVAVFMHMAWERLALIYAIILPPLLVLVFVALMVS
                ESNYVLFTRLTFFGGGE"
gene            5477. .8038
                /gene="SMc01987"
CDS             5477. .8038
                /gene="SMc01987"
                /function="small molecule metabolism"
                /note="Product confidence : putative
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11

alignment_scores:
           Quality:  127.00            Length:    230
             Ratio:  1.187               Gaps:     13
 Percent Similarity:  46.522   Percent Identity: 23.913

alignment_block:
 US-09-674-779-2 x SME591791/rev   ..

 Align seg 1/1  to reverse of: SME591791  from: 1  to: 340900

      20 CysSerAlaProIleProThrAsnProGln..................... 29
         |||||||||               ||||||
  299697 TGCTCGGCGGTCGCTCGAAAGGCGCCCCAGGTCCTGGCGTCCATCGCGCT 299648

      30 .ValSerProIleLysThrProSerValLeuIleThrLysAspLysIleG 46
          |||       :::   ||||||::::::            :::      |
  299647 GGTTTGCTCCCTTGTCACACCCGGCATGGCGCCTCCGGTCGAGGCCGCCG 299598

      46 lyAspHisHisThrHisGluHisAspGluSerValSerHisValGlyLeu 62
         ||:::   :::|||                                 |||
  299597 GCCAGACACGCACC..............................CTC 299581

      63 GlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGl 79
         :::   :::|||               :::|||         |||   ::
  299580 AAGCTCTATTTC..............ATCCACACCAAGGAAAAGGCGCA 299546

      79 uValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr...L 95
         :::::::                     :::|||   |||:::|||
  299545 GATCACC..................TACAAGCGCAACGGTCGCTACGATC 299514

      95 euProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys 111
               ::::::|||:::         |||:::|||:::
  299513 AGAAGGGTCTGCAGCAGATCAATCGCTTCCTGCGCGACTGGCGGCGA... 299467

     112 GlyHisGluProTyrGlnLeuProProGluHisLeu.........TrpGl 125
            :::|||||   ::::::   |||:::   |||         |||
  299466 ...AACGAGCCGACCAAAATGGACCCGCGCCTGCTCGATCTCGTCTGG.. 299422

     125 yGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyI 142
                       ::::::|||||         ||||||
  299421 ...................GAAGTTTACCAGAAGAGCGGTTCTCGCGACT 299391

     142 leLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu 158
         :::           ::::::   |||:::||||||:::|||
  299390 ACATC............CATGTCGTTTCCGCCTATCGCTCCCCCGCCACG 299353

     159 Asn..............GlnCysAlaGlyGlyAlaAlaMetSerLysHi 170
         |||             :::      |||   |||      |||:::||
  299352 AACGGCATGCTGAGGTCGCGGTCGAAGGGCGTCGCGAAGAAGAGCCAGCA 299303
```

```
    170 sLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysS 187
        |:::   :::::::|||:::|||:::::::::|||||||:::::::::|||:
  299302 CATGCTCGGCAAGGCGATGGACTTCTATATACCGGACGTGAAGCTCAAGA 299253

    187 erGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGlu 203
        ::    |||   |||:::      ::::::
  299252 CG......CTGCGCGAGGTCGGCATGAAGTTC..................... 299227

    204 HisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAla.. 219
                |||      |||:::|||   |||   |||::::::
  299226 ............CAGGTGGGCGGCGTCGGCTATTATCCGACTTCCGGTTC 299189

    220 .......IleHisLeuAspThrGlnGlyPheArgLysTrp 230
              :::|||:::|||:::   |||   |||   |||
  299188 GCCCTTCGTCCATATGGACGTCGGTGGCGTGCGAGCCTGG 299149
```

```
seq_name: gb_ba:AP003009

seq_documentation_block:
LOCUS       AP003009              339681 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 16/21.
ACCESSION   AP003009 BA000012
VERSION     AP003009.2  GI:14026063
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 339681)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994984.
FEATURES             Location/Qualifiers
     source          1. .339681
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
     gene            complement(82. .420)
                     /gene="mll6350"
     CDS             complement(82. .420)
                     /gene="mll6350"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52662.1"
                     /db_xref="GI:14026064"
                     /translation="MPLLAFDLLARIALPKVGLPLERVNLEACSLTAWPHDGLRAEGQ
                     PVHVKSWQCRLWRSLLVHRHPVLNEMRSIENMVRTILREACINWARGREPPSPVVCRS
                     WPALTPRCWK"
     gene            447. .593
                     /gene="msr6351"
     CDS             447. .593
                     /gene="msr6351"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52663.1"
                     /db_xref="GI:14026065"
                     /translation="MHALVHAADIQDCDGGVMVMAKLFGLDPFLFKLYAGGGHSRGRS
                     SKRG"
     gene            complement(733. .1464)
                     /gene="mll6352"
     CDS             complement(733. .1464)
                     /gene="mll6352"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52664.1"
                     /db_xref="GI:14026066"
                     /translation="MRTWAMSEVARDPLYRCHRFPAEIIAHAVWLYFRFPLSLRMVED
                     MLAARGIIVTHQTIRSWAEKFGRHFAREIKRRSAGCLGDKWHLDECVVAINGKKQWLW
                     RAVDQDGLVLEVLVQSRRNAKAAKRLMRKLLKAQGRTPRVMITDKLRSYDAARRDLMP
                     GVEHRSHKGLNNRAENSHQPTRRRERTMKRFKSARQLQRFVSIHDPIANLFHFPRNTL
                     SSAQHRDLRNAAMQIWNKIACLAAA"
     gene            complement(1080. .1748)
                     /gene="mll6353"
     CDS             complement(1080. .1748)
                     /gene="mll6353"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52665.1"
                     /db_xref="GI:14026067"
                     /translation="MVEACATYKNHRFPIEIVAHAVWLYYRFGLSLRDVEEMLLERGI
                     VVSYETIRRWGKKHSPDYARRLRRKAPSKDDVWHLDEVAGHRKVNRRLMKMCEHGPCQ
                     KLLVIRFIVATAFPPRSLRTRYGSTFAFRSACAWLRTCWRPAASSSRIRPFEAGRRNS
                     GGISPGRSSDGQPAAWATNGISTNVWWPSMARSSGSGVPSIRTAWSSKCWCKAAEMPR
                     RQSV"
     gene            complement(1766. .2200)
                     /gene="mll6354"
     CDS             complement(1766. .2200)
                     /gene="mll6354"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52666.1"
                     /db_xref="GI:14026068"
                     /translation="MVARHATIAAEAVERTRTPPSRVRRQAEHAPVSAKELAQALPED
                     AWHMVTWREGSKASLASRFATVRVRPAHRDYWRSIPHAEEWFLIEWPDGEVEPTKYWL
                     STLPGKTRLADLVDQAKMRWRIERDYQELKQGCCHIDFGLTV"
     gene            complement(1939. .2931)
                     /gene="mll6355"
     CDS             complement(1939. .2931)
                     /gene="mll6355"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52667.1"
                     /db_xref="GI:14026069"
                     /translation="MLHMEGAMDLQDGIETSESRFAAYVETLASVLGHADRVAPLKAY
                     CTGLLLPRERKSVEPMAARVEPGRVQAAHQSLHHFVAKADWSDDAVLSRVREQVLPAL
                     ERQGQIRAWIVDDTGFPKKGKHSVGVARQYCGQLGKQDNCQVAVSLSVATEQASLPIA
                     YQLYLPESWANDPDRRAKTGVPENVVFRTKPEIALAQIRAALEAGVSRGVVLADAGYG
                     NDTAFRTGLTKVGLTYVVGVQSSIRLWSPGTQPLPPKPWSGRGPRPRAFGGRLNMRRS
                     QPRNWRKRCPRTPGTWLLGARAPRRRLPHALPPCGFGRPIATTGAPSPTLRNGS"
     gene            complement(3398. .3616)
                     /gene="msl6356"
     CDS             complement(3398. .3616)
                     /gene="msl6356"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52668.1"
                     /db_xref="GI:14026070"
                     /translation="MLSGMPPCRLLRTMMMAGATIDLLAGQAFLVAGFQIETRTARHH
                     RSAAKRDPRSPGLRNLTNSARPNADRSL"
     gene            4196. .6634
                     /gene="mlr6358"
     CDS             4196. .6634
```

```
                     /gene="mlr6358"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52669.1"
                     /db_xref="GI:14026071"
                     /translation="MADTSAPIHRQRVLVYIGDEKMTGIGRSGKSRIGDAGAGSSSSR
                     SQSGLASGEGSQGFGSVLERMRSEGGKGSPSAGAARGAHDDLGGKSSGPSVPPHRIPA
                     AAPRAAAGPRAAGSRRRSVSSTSDGEALAPVASQPSTSGVTVPSSSSGEVINETKRRM
                     CSLIDEIQVLEEKSLHSGKPMEGFRDELDQLANAGAELLECYASLPKDEARSTLSDDD
                     VRHYRADVIHRADSANELAHTIFKKEEQNRKRMEDILHNLVKSNATLDQLSRSAPSMA
                     KYPVACVERWDKKVQRHELMRAVCVATADLPSSTAEMRREEEAALRLHSGWVVNSKIM
                     QLQSRINLAQVNMAPQAIMLEAPIRELLIGDNGQTRLLGNFIGEVFPAFVSTAEAVLR
                     KEGRPLDAEHCAVLEGVVERLADFASASHSVVDRLRDHQAGAGLPLELLTRIVDDAWI
                     ASDEVTDLLALQPKPPAPDLTADAAAAMPTGTAGKTMVAESTTAPRKKGKDKSKQAAG
                     AGSSAAGRPEPQVAAAVSDPAPAAKVLVRTDLGTKKLVSAEEAQNSASSAAARLAILQ
                     NPTTGENTKSLVARSTELLKFDLPAQQSLISRARSGKPEDVEHAVGLVVQRLQTQAVE
                     MEPCLPALEASRLSKEGKMDDVHRLTVQFRDMLANVRGLAKAWDEQKPDMIMDCMKRY
                     PFPSQNHLEQLRSARELTPADPPCALNPEPSNLFEIRLQPKALSNRTLPSPMWVHIHT
                     TRKIYAQQLATLDSAEFAACHVKSNEQRGHNRQWQTSRAKKGYDNAIHRGKLTPAFCK
                     SLLSDRADMRNGPA"
gene                 complement(6985. .7716)
                     /gene="mll6359"
CDS                  complement(6985. .7716)
                     /gene="mll6359"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52670.1"
                     /db_xref="GI:14026072"
                     /translation="MRTWAMSEVARDPLYRCHRFPAEIIAHAVWLYFRFPLSLRMVED
                     MLAARGIIVTHQTIRSWAEKFGRHFAREIKRRSAGCLGDKWHLDECVVAINGKKQWLW
                     RAVDQDGLVLEVLVQSRRNAKAAKCLMRKLLKAQGRTPRVMITDKLRSYDAARRDLMP
                     GVEHRSHKGLNNRAENSHQPTRRRERTMKRFKSARQLQRFVSIHDPIANLFHFPRNTL
                     SSAQHRDLRNAAMQIWNKIACLAAA"
gene                 8086. .17256
                     /gene="mlr6361"
CDS                  8086. .17256
                     /gene="mlr6361"
                     /note="similar to shikimate kinase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52671.1"
                     /db_xref="GI:14026073"
                     /translation="MTGVNRSGRSRTEKRVGRTGEASSSRSQSSLAPGAGSQSFDTVV
                     EQMRSVEINRRTPAVPMERGADDALHDRPSGPSVGIHGTLDAARHAAAAPRAAGPRRR
                     SVPSMLEGETLAPVASQPPTSGVAAPICLSGDEIKEKRRMDSLLTLVGMRFNAARQDP
                     GDEKLLDQVINAGSAALHHYKSLSPALARTIVDDNQLITCRYLVWDAVQEYSRLAEPT
                     IHKLQKKKAEEEAIGLTQSTATPESVVEAHAACVKWWEKKVGHCERGRAAWRATASLP
                     SATAKMRQEEEATLRLATGSVMHSKLVYLQSRIDLARAKIEPQASTFDAPVREILMVE
                     NGRVRLLEAFSGDVLPAFLSTQNAVLSSGGHPLDADHCAVLEGVIERLSEFASELQVI
                     RTNLNDHQTSADLPLELLSQIVEGAWIVGHEVMHLLELQPKPPAVALPPDAGAAIPAE
                     TAGKSRVAKGTLALQPKAPALIASPADAGATIQIDTVGKSPVAKGTLALQPNWPANIA
                     PPADMGAGIPADTAGKAAVAGAAARRKRKGKGKPAVGAGSSAPGRASQVAAVDSDTAP
                     ADKVLVLSDMGTKLVSAEEARASASAGAQSAKQAPPSMEALTQERQVKPEDARYVAET
                     VVERLQTQPAEMRACPAALDEPRRPGLLAPAQVPEVQDKTVRLKGMLNQAQELAKALN
                     SQSSHAPGQSSQSSDAGVERTHSGPQDRMSYARVVRGADEALGDKSADYSIAPRGALA
                     GSRGATSGPRHGSPQGRHGEAMGRQESSLRLERSWLGDQEPSRSQPDTLARHDQSQSR
                     PVAGPERRRFDHVAAQSFDDALLNDALHKIFFAHDINEFSYAVVENDKRLQSQAHRAT
                     FLRKAASEFTRHFVPRWEDDVRAGNTWGFATSCNATSREPGGQAGMEACRAMAAQVSR
                     LGRALNDVESKTLSLFVLSFSRHPRLAECRNGMITIAKLFHEQPGALSELNGQSLALL
                     VNGFSKWPEQENARKATVAVAAEIRRRPAGLSDFDPQNLGNVVNGLSKWPLEEECCGV
                     IDAIAGKVCGRAGEKAGLSGFNSQAFANLVNGFSKAREPGNSHQATIVIADAVCRLAD
                     EIVIADSIGRLADEKAGRTGWYPQALANLVNGFSKWPLDCRGAIVAIARQVAGGAGKL
                     PVFDAQALANLVNGFSKWPEDCRGATIAIAAEVCRLADLPKAGLSGFDPQHLANLVNG
                     FSKLQAAECSRATVAIAGEVCHRAGRRKGLSDFDVQHLANLVKGFSKWPQEVNLGYAT
                     NVVAGEVLRCAEPLSEFLPLHLASLVNGFSKWPRWENSHQATVAIAGEVRHRADQLSG
                     FEPQELANLVNGFSKWPEDCRGAIVAIARAVPRRAGRLSGFNGQDLANLVNGFSKWPL
                     DCREAMVAIARAVPRRAGRLSGFNGQDLANLVNGFSKWPQEENSRQATVAIAGEVLRS
                     TNPPPDFSSQYLANLVNGFSKWPEEEALRQAAAAIAGEVIRSANRLPDFTSQHLANLV
                     NGFSKWPQEEKSRQATAAIAGEVLRGAERLSGVNPQQLANLLNGFDRFAEEEACSQAI
                     LEIARRLGQAGQPFRHFAAPGLSSIANSLARGILRSEDAGEIAEAALLKDRLHKLAHY
                     LHYASDRLEEADAVGVTSILKALAKAQLYDDLGSLAGAGLNRLAELHRAPGFAHENNL
```

```
                                                ETMGNLCAALLPLARSPRKQLLWHRRPALNLLNDIQPIVEHKIEAHLRASDAERARGP

alignment_scores:
            Quality:  112.00              Length:    264
              Ratio:  0.896                Gaps:     14
 Percent Similarity:  47.348   Percent Identity: 22.348

alignment_block:
 US-09-674-779-2 x AP003009   ..

 Align seg 1/1  to: AP003009  from: 1  to: 339681

      5 AsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuValAlaCys.. 20
        ||||||      :::         :::::::::   :::   ||||||
 134211 AACCAAGATGCTGTTCATCGACGGGTCGCAGGACGGATGCCCGCCTGCCT 134260

     21 ....SerAlaProIleProThrAsnProGlnValSerProIleLysThrP 36
            :::         |||         :::   :::|||:::      |
 134261 GGTGAAACCTGCGTGTCCAGGGTTCGAAAAGAAGGCTCCGCTT......C 134304

     36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGlu 52
        ||:::::      ::::::   |||      :::         |||:::
 134305 CGGCGACGACGTCAAATCGTCCGAAACAAGAACAATGTGCAAGGCATGAC 134354

     53 HisAsp.GluSerValSerHisValGlyLeuGlnAlaHisPheGluThrT 69
        :::||| ||||||   :::::::::         :::|||      ||||
 134355 CGTGATTGAAAGCGAACAGAACGCGCCCGGCGCCTCCCATCGTTGGACAT 134404

     69 rpLeu..................................................... 70
        |||||
 134405 GGTTGCCGGCCATTGTCGTCGCGTTCCTTTGCCTGTGCGTCACCAGCGCG 134454

     71 ..........................GlnMetHisHisAlaThrLy 77
                                  :::::::|||   :::   ||
 134455 TTCGCCGAGACGCGTGCCTTGAAGATCCAGCATCTGCATACGGGCGAGAA 134504

     77 sGlnGluValVal......ArgTyrGlnAlaTyrLeuGlnSerArgLeuG 92
        |   |||:::|||      |||         |||   |||:::   |||
 134505 GGCGGAGATCGTCTTCAAGCGCAACGGACGCTATGATCAGGCGGGATTG. 134553

     92 lyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrp 108
                        :::::::::            |||:::|||
 134554 ...................AAGAAGATCGATTTCATGCTGCGCGACTGG 134583

    109 GlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGl 125
        :::         :::|||||||   ::::::   |||::   |||
 134584 CGCCGC......AACGAGCCGACCAGGATGGATCCGCGCCTGCTC..... 134622

    125 yGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyI 142
         :::::|||               :::|||      ::::::   |||
 134623 .GATCTCGTC..............TGGCAGGCCTATCGCGCCAGCGGC. 134655

    142 leLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu 158
               |||      ::::::   |||:::||||||:::|||
 134656 ..TCGAGCGCCTATATCCATGTCGTCAGCGCCTATCGCTCGCCGGCGACC 134703

    159 Asn..............GlnCysAlaGlyGlyAlaAlaMetSerLysHi 170
        |||               :::      |||   |||      |||:::||
 134704 AACGCCATGCTGCGCAGCCGCTCGAAGGGCGTGGCGCGGGAAAGCCAGCA 134753

    170 sLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysS 187
        |:::::::::   |||:::|||:::::::|||||:::   :::|||:
 134754 CATGGTCGGCCGGGCCATGGACTTCTTCCTGCCGGACGTGCCGCTGAAGA 134803

    187 erGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGlu 203
        ::      |||   ::::::      ::::::
 134804 AG......CTGCGCGACATCGGCCTCAAGATG..................... 134829

    204 HisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAla.. 219
                |||:::   |||:::|||   |||   |||::::::
 134830 ............CAGGGCGGTGGTGTTGGTTATTATCCGACCTCCGGCTC 134867

    220 .......IleHisLeuAspThrGlnGlyPheArgLysTrp 230
```

```
                ||||||:::|||:::   :::   |||   |||
  134868 GCCTTTCATTCACATGGACGTCGGCAATGTGCGGCATTGG 134907

seq_name: gb_htg:AC046147

seq_documentation_block:
LOCUS       AC046147               120405 bp    DNA     linear   HTG 16-OCT-2001
DEFINITION  Mus musculus chromosome 12 clone RP23-265L9, *** SEQUENCING IN
            PROGRESS ***, 38 unordered pieces.
ACCESSION   AC046147
VERSION     AC046147.5  GI:16118086
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 120405)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 120405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 14, 2001 this sequence version replaced gi:11094635.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MADW
                Center clone name: RP23-265L9
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 147066 bases at least Q40
                Consensus quality: 188973 bases at least Q30
                Consensus quality: 206320 bases at least Q20
                Estimated insert size: 186499; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 38 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5655: contig of 5655 bp in length
            *     5656     5755: gap of unknown length
            *     5756    10324: contig of 4569 bp in length
            *    10325    10424: gap of unknown length
            *    10425    14993: contig of 4569 bp in length
            *    14994    15093: gap of unknown length
            *    15094    20176: contig of 5083 bp in length
            *    20177    20276: gap of unknown length
            *    20277    23825: contig of 3549 bp in length
            *    23826    23925: gap of unknown length
            *    23926    28679: contig of 4754 bp in length
            *    28680    28779: gap of unknown length
            *    28780    31896: contig of 3117 bp in length
            *    31897    31996: gap of unknown length
            *    31997    35381: contig of 3385 bp in length
            *    35382    35481: gap of unknown length
            *    35482    39457: contig of 3976 bp in length
            *    39458    39557: gap of unknown length
            *    39558    43791: contig of 4234 bp in length
            *    43792    43891: gap of unknown length
            *    43892    46680: contig of 2789 bp in length
            *    46681    46780: gap of unknown length
            *    46781    49552: contig of 2772 bp in length
            *    49553    49652: gap of unknown length
            *    49653    52860: contig of 3208 bp in length
            *    52861    52960: gap of unknown length
            *    52961    55985: contig of 3025 bp in length
            *    55986    56085: gap of unknown length
            *    56086    58677: contig of 2592 bp in length
            *    58678    58777: gap of unknown length
            *    58778    62289: contig of 3512 bp in length
            *    62290    62389: gap of unknown length
            *    62390    65117: contig of 2728 bp in length
            *    65118    65217: gap of unknown length
            *    65218    67647: contig of 2430 bp in length
            *    67648    67747: gap of unknown length
            *    67748    71042: contig of 3295 bp in length
            *    71043    71142: gap of unknown length
            *    71143    73664: contig of 2522 bp in length
            *    73665    73764: gap of unknown length
            *    73765    76509: contig of 2745 bp in length
            *    76510    76609: gap of unknown length
            *    76610    79272: contig of 2663 bp in length
            *    79273    79372: gap of unknown length
            *    79373    81777: contig of 2405 bp in length
            *    81778    81877: gap of unknown length
            *    81878    84300: contig of 2423 bp in length
            *    84301    84400: gap of unknown length
            *    84401    87125: contig of 2725 bp in length
            *    87126    87225: gap of unknown length
            *    87226    89379: contig of 2154 bp in length
            *    89380    89479: gap of unknown length
            *    89480    92629: contig of 3150 bp in length
            *    92630    92729: gap of unknown length
            *    92730    95493: contig of 2764 bp in length
            *    95494    95593: gap of unknown length
            *    95594    98776: contig of 3183 bp in length
            *    98777    98876: gap of unknown length
            *    98877   101072: contig of 2196 bp in length
            *   101073   101172: gap of unknown length
            *   101173   103468: contig of 2296 bp in length
            *   103469   103568: gap of unknown length
            *   103569   106241: contig of 2673 bp in length
            *   106242   106341: gap of unknown length
            *   106342   108452: contig of 2111 bp in length
            *   108453   108552: gap of unknown length
            *   108553   110640: contig of 2088 bp in length
            *   110641   110740: gap of unknown length
            *   110741   113610: contig of 2870 bp in length
            *   113611   113710: gap of unknown length
            *   113711   116067: contig of 2357 bp in length
            *   116068   116167: gap of unknown length
            *   116168   118270: contig of 2103 bp in length
            *   118271   118370: gap of unknown length
            *   118371   120405: contig of 2035 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .120405
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /clone="RP23-265L9"
BASE COUNT    35100 a  22309 c  23576 g  35674 t   3746 others
ORIGIN


alignment_scores:
           Quality:  110.00            Length:    208
             Ratio:  1.310               Gaps:     10
 Percent Similarity:  40.385   Percent Identity: 23.558

alignment_block:
 US-09-674-779-2 x AC046147/rev   ..

 Align seg 1/1  to reverse of: AC046147  from: 1  to: 120405

      48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
         ||||||||||||   |||   ::::::   :::|||:::   ::::::::
   91957 CACCACACACACACACACACACACACACACACACACACACACACATACACAC 91908

      64 aHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValV 81
         :|||                    ||||||
   91907 ACAC.................ACACACCACATA..................... 91892

      81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
                                         |||||||||
   91891 .....................................TATCTGCCC... 91883

      98 MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114

   91883 ................................................. 91883

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrL 131

   91883 ................................................. 91883

     131 euHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThr 147
                |||||||||:::   |||   :::||||||||||:::
   91882 ...........CAGGATTTGAGCTAGAGACCCATGCTACCAGCCAGC... 91847

     148 GlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGl 164
         :::|||                      :::::::|||
   91846 ...CTGAGGTTGTACGTGGGTCTGGGGACCACATCCTCATGC........ 91808

     164 yAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp...IleTrpV 180
                :::|||||||||||||:::   |||||||:::      ||||
   91807 ....TTGCCTGACAAGCACCTTACCCACTGAGCTATCTCCCCAGCCTGGG 91762

     180 alProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeu......Gln 194
         |||||      :::::::::   ::::::      :::|||      :::
   91761 TACCACACACTAAATTAGAGTTTGAGACTTAGTTAAGCTTATGACAAGAG 91712

     195 AsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGl 211
         :::|||      :::   |||||||||      :::|||   |||   ::
   91711 GACAGGAGTGGGGAGTGTTGGCTGGAAACACAGCAAAATCTGAATGTGGC 91662

     211 yLeuGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe. 227
         :   |||                    :::|||   :::|||   |||
   91661 TTCAGGT.................TGACAACTACATTCACAACCTTTTT 91630

     228 ........ArgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCys 241
                 :::|||         ||||||:::|||   :::   |||
   91629 CAAGCCGGGCATCATGG......GTATTCTCAGATACTTATGGGCATTGT 91586

     242 ArgHisValLeuProLysAsnLys 249
         |||::::::   |||:::|||
   91585 ACTCACACTATCCTGAAAACAAAG 91562

seq_name: gb_htg:AC073821
```

```
seq_documentation_block:
LOCUS       AC073821              145347 bp    DNA     linear   HTG 02-SEP-2000
DEFINITION  Mus musculus clone RP23-79P7, WORKING DRAFT SEQUENCE, 10 ordered
            pieces.
ACCESSION   AC073821
VERSION     AC073821.2  GI:9964818
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 145347)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 145347)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 2, 2000 this sequence version replaced gi:8810438.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1767575
            Center clone name: RPCI-23_79P7
            -------
            Summary Statistics
            Consensus quality: 138992 bases at least Q40
            Consensus quality: 143251 bases at least Q30
            Consensus quality: 144092 bases at least Q20
            Estimated insert size: 150000; agarose-fp estimation
            Estimated insert size: 144947; sum-of-contigs estimation
            Quality coverage:  9.13 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.45 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    20853: contig of 20853 bp in length
            *    20854    20953: gap of unknown length
            *    20954    54936: contig of 33983 bp in length
            *    54937    55036: gap of unknown length
            *    55037    75800: contig of 20764 bp in length
            *    75801    75900: gap of unknown length
            *    75901    79363: contig of 3463 bp in length
            *    79364    79463: gap of unknown length
            *    79464    82637: contig of 3174 bp in length
            *    82638    82737: gap of unknown length
            *    82738   107168: contig of 24431 bp in length
            *   107169   107268: gap of unknown length
            *   107269   109162: contig of 1894 bp in length
            *   109163   109262: gap of unknown length
            *   109263   136238: contig of 26976 bp in length
            *   136239   136338: gap of unknown length
            *   136339   137999: contig of 1661 bp in length
            *   138000   138099: gap of unknown length
            *   138100   145347: contig of 7248 bp in length.
FEATURES             Location/Qualifiers
     source          1. .145347
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-79P7"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    37072 a  35595 c  34914 g  36864 t    902 others
```

```
ORIGIN

alignment_scores:
            Quality:  110.00              Length:    243
              Ratio:  0.932                Gaps:     16
 Percent Similarity:  48.560   Percent Identity: 26.337

alignment_block:
 US-09-674-779-2 x AC073821/rev   ..

 Align seg 1/1  to reverse of: AC073821  from: 1  to: 145347

      13 IleSerSerMetLeuValAlaCysSerAlaProIleProThrAsnProGl 29
         :::|||||||:::   |||:::          |||:::   ::::::|||||
    6874 CTGTCTTCAGTGTATGTCTGTCTGCGACATCCACTCCAAGTTCAGCCCCA 6825

      29 nValSerProIleLysThrProSerValLeuIleThrLysAspLysIleG 46
         ||||          :::   ::::::|||
    6824 AGTCTTTTCTACATGGGTCACTGCCATTCTCTGCATA..............  6788

      46 lyAspHisHisThrHisGluHisAspGluSerValSerHisValGlyLeu 62
            ||||||||||||   |||   ::::::   :::|||:::
    6787 .....CACCACACACACACACACACACACACACACACACACACACACACA 6743

      63 GlnAlaHisPheGluThrTrp......LeuGlnMetHisHisAlaThrLy 77
         ::::::|||   :::|||         :::::::::||||||
    6742 CACACACACACACACTCATGCAGTTGTGAACTTTCATCACAGAAAAGG 6693

      77 sGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnT 94
          ::::::              |||   ::::::||||||   |||:::
    6692 AGAGCAA............GGGGCTCTCATCCACTCCCGCTGTGGCTCC. 6656

      94 yrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAla 110
                |||:::::::::|||:::
    6655 ........CCACTCAAGTCACTGGTC.....................CTT 6635

     111 CysGlyHisGluProTyrGlnLeu.ProProGluHisLeuTrpGlyGlnI 127
         |||   |||:::|||   |||       |||   |||||||||
    6634 TGTGTCCACCACCCTTGACAGACATTATCCTCCTCACCTTTGGCGG.... 6589

     127 leValProThrLeuHis.........LeuTyrGlnAspLeuLys...... 138
           :::|||      |||         |||:::|||   ||||||
    6588 ..ACTCCCCCACAGCATTTGGACTTTCTGCACCAGCACTTAAAAAAAAAA 6541

     139 ...............SerArgGlyIleLeuProAlaAsn......ThrGl 148
                       |||      :::|||      |||      |||::
    6540 AGGTCTCACTGTGTCTCCTCGACTGTTCTGGAACTCAATCTGTGGACCAA 6491

     148 nIleArgSer.ValTyrArgAsnPro.Glu..................... 157
         ::::   ||| :::   |||:::|||
    6490 GCTGGCCTCAGACTAACAGAGAGCCACACTGACCTCGGGCTGGTCCTCTG 6441

     158 ..........LeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
                 |||:::   ||||||   :::::::::   :::   |||
    6440 GGTGTTGGGATTAAAAGCATGCGCCACCAGCGTGTGCCACCAAGGTCACT 6391

     171 euThrAsnSerAlaIleAspIle..............TrpValProAsp 182
            ::::::::::::   |||              :::      |||
    6390 GCCACCAAGCCTCCCTCGGCATCCCCTGAGTGTTTTCTTCCCCTCTGAT 6341

     183 LeuGluIleLysSerGlnAla.....................LeuTyrGl 192
         ::::::   |||   ||||||                    ||||||
    6340 TTCTCACCAAAATGTCAAGCTTCTTCTTATTTTATTTTTATTCTTTATTT 6291

     192 uLeu......GlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluA 207
          |||      |||:::|||      |||   |||::::::   ::::::
    6290 ATTGGTTTTTCAAGAGAGGGCTTTTCAGCCTTGGATGTCCTGGAACTCAG 6241

     207 snGlnAsnPheGlyLeuGlyLeu 214
         :::      ||||||   |||
    6240 TAGACCCGACTGGCCTTGAACTC 6218
```

```
seq_name: gb_htg:AC073767

seq_documentation_block:
LOCUS       AC073767              201377 bp    DNA     linear   HTG 18-JUL-2000
DEFINITION  Mus musculus clone RP23-359H6, WORKING DRAFT SEQUENCE, 13 ordered
            pieces.
ACCESSION   AC073767
VERSION     AC073767.2  GI:9256787
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201377)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 201377)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:8810384.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1874902
            Center clone name: RPCI-23_359H6
            -------
            Summary Statistics
            Consensus quality: 189664 bases at least Q40
            Consensus quality: 197375 bases at least Q30
            Consensus quality: 198999 bases at least Q20
            Estimated insert size: 176970; agarose-fp estimation
            Estimated insert size: 200827; sum-of-contigs estimation
            Quality coverage:  8.42 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.42 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     9998: contig of 9998 bp in length
            *     9999    10098: gap of unknown length
            *    10099    17043: contig of 6945 bp in length
            *    17044    17143: gap of unknown length
            *    17144    40422: contig of 23279 bp in length
            *    40423    40522: gap of unknown length
            *    40523    70835: contig of 30313 bp in length
            *    70836    70935: gap of unknown length
            *    70936    82602: contig of 11667 bp in length
            *    82603    82702: gap of unknown length
            *    82703    97050: contig of 14348 bp in length
            *    97051    97150: gap of unknown length
            *    97151   116905: contig of 19755 bp in length
            *   116906   117005: gap of unknown length
            *   117006   132589: contig of 15584 bp in length
            *   132590   132689: gap of unknown length
            *   132690   135229: contig of 2540 bp in length
            *   135230   135329: gap of unknown length
            *   135330   147912: contig of 12583 bp in length
            *   147913   148012: gap of unknown length
            *   148013   159289: contig of 11277 bp in length
            *   159290   159389: gap of unknown length
            *   159390   175515: contig of 16126 bp in length
            *   175516   175615: gap of unknown length
```

```
          *   175616  201377: contig of 25762 bp in length.
FEATURES             Location/Qualifiers
     source          1. .201377
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-359H6"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    49152 a  48199 c  50595 g  52229 t   1202 others
ORIGIN


alignment_scores:
            Quality:  110.00              Length:    243
              Ratio:  0.932                Gaps:     16
 Percent Similarity:  48.560    Percent Identity: 26.337

alignment_block:
 US-09-674-779-2 x AC073767   ..

 Align seg 1/1  to: AC073767  from: 1  to: 201377

      13 IleSerSerMetLeuValAlaCysSerAlaProIleProThrAsnProGl 29
         :::||||||:::   |||:::         |||:::   :::::|||||
   87509 CTGTCTTCAGTGTATGTCTGTCTGCGACATCCACTCCAAGTTCAGCCCCA 87558

      29 nValSerProIleLysThrProSerValLeuIleThrLysAspLysIleG 46
         ||||            :::   ::::::|||
   87559 AGTCTTTTCTACATGGGTCACTGCCATTCTCTGCATA..............  87595

      46 lyAspHisHisThrHisGluHisAspGluSerValSerHisValGlyLeu 62
            ||||||||||||   |||   ::::::   :::|||:::
   87596 .....CACCACACACACACACACACACACACACACACACACACACACACA 87640

      63 GlnAlaHisPheGluThrTrp......LeuGlnMetHisHisAlaThrLy 77
         ::::::|||   :::|||         :::::::::||||||
   87641 CACACACACACACACTCATGCAGTTGTGAACTTTCATCACAGAAAAGG 87690

      77 sGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnT 94
         ::::::              |||   ::::::||||||   |||:::
   87691 AGAGCAA............GGGGCTCTCATCCACTCCCGCTGTGGCTCC. 87727

      94 yrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAla 110
               |||::::::::::|||:::
   87728 ........CCACTCAAGTCACTGGTC.....................CTT 87748

     111 CysGlyHisGluProTyrGlnLeu.ProProGluHisLeuTrpGlyGlnI 127
         |||  |||:::|||  |||       |||  |||||||||
   87749 TGTGTCCACCACCCTTGACAGACATTATCCTCCTCACCTTTGGCGG.... 87794

     127 leValProThrLeuHis.........LeuTyrGlnAspLeuLys...... 138
           :::|||      |||         |||:::|||   ||||||
   87795 ..ACTCCCCCACAGCATTTGGACTTTCTGCACCAGCACTTAAAAAAAAAA 87842

     139 ..............SerArgGlyIleLeuProAlaAsn......ThrGl 148
                      |||      :::|||      |||      |||::
   87843 AGGTCTCACTGTGTCTCCTCGACTGTTCTGGAACTCAATCTGTGGACCAA 87892

     148 nIleArgSer.ValTyrArgAsnPro.Glu.................... 157
         ::::   ||| :::   |||:::|||
   87893 GCTGGCCTCAGACTAACAGAGAGCCACACTGACCTCGGGCTGGTCCTCTG 87942

     158 ..........LeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
                 |||:::   ||||||   :::::::::   :::   |||
   87943 GGTGTTGGGATTAAAAGCATGCGCCACCAGCGTGTGCCACCAAGGTCACT 87992

     171 euThrAsnSerAlaIleAspIle..............TrpValProAsp 182
            ::::::::::::   |||               :::      |||
   87993 GCCACCAAGCCTCCCTCGGCATCCCCTGAGTGTTTTTCTTCCCCTCTGAT 88042

     183 LeuGluIleLysSerGlnAla...................LeuTyrGl 192
         ::::::   |||   ||||||                  ||||||
   88043 TTCTCACCAAAATGTCAAGCTTCTTCTTATTTTATTTTTATTCTTTATTT 88092

     192 uLeu......GlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluA 207
            |||      |||:::|||      |||   |||:::::::   ::::::
   88093 ATTGGTTTTTCAAGAGAGGGCTTTTCAGCCTTGGATGTCCTGGAACTCAG 88142

     207 snGlnAsnPheGlyLeuGlyLeu 214
         :::      ||||||   |||
   88143 TAGACCCGACTGGCCTTGAACTC 88165

seq_name: gb_ba:AP003000

seq_documentation_block:
LOCUS       AP003000              348077 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 7/21.
ACCESSION   AP003000 BA000012
VERSION     AP003000.2  GI:14023048
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 348077)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994975.
FEATURES             Location/Qualifiers
     source          1. .348077
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
     gene            complement(169. .1065)
                     /gene="mll2551"
     CDS             complement(169. .1065)
                     /gene="mll2551"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transcriptional regulator"
                     /protein_id="BAB49656.1"
                     /db_xref="GI:14023049"
                     /translation="MALDWDKLRVFHAAAEAGSFTHAAETLHLSQSAISRQVSALEHD
                     VGVPLFNRHARGLVLTEQGEMLFRTAHDVLMKLETIKSRLTETKDRPSGVLRVTTTVG
                     LGAGWLTERVQEFIELYPEISLQLILANEELDLTMR@ADCAIRLRQPQQPDLIQRRLF
                     TVHFHLYAAPSYVAKHGKPASVSELRNHRIVTFGLPVPSHLSELNWLETVGDFEGGQR
                     VPSLQINDILSIKRAVQGGAGIAMLPDYVINKDSGLVQLLPETEVPSFDTYFAYPDAM
                     KNQAKLHVFRDFIIAKARSWSF"
     gene            complement(1156. .2130)
                     /gene="mll2552"
     CDS             complement(1156. .2130)
                     /gene="mll2552"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin reductase"
                     /protein_id="BAB49657.1"
                     /db_xref="GI:14023050"
                     /translation="MTTKHAPVLIIGSGPAGYTAAVYAARAMLKPMLVAGLQQGGQLM
                     ITTDVENYPGFADPIQGPWLMEQMMKQAEHVGTDIINDIITEVDLNVRPFRAKGDSGT
                     TYTADALIIATGAQAKWLGIPTEQDFMGFGVSACATCDGFFYRGKDVAVVGGGNSAVE
                     EALYLSNLAKSVTVIHRRSDFRAERILRERLLQKDNVRVIWDTVVDEITGRPGKAPLP
```

```
                PSVEGLKLKHAVTGAETHLKVDGVFVAIGHAPAVELFVGKLKQKPNGYLWTAPNSTRT
                DVPGVFAAGDVTDDVYRQAVTAAGLGCMAALEAEKYLAGIEVHREAAE"
     gene       2378. .2857
                /gene="mlr2554"
     CDS        2378. .2857
                /gene="mlr2554"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="BAB49658.1"
                /db_xref="GI:14023051"
                /translation="MPLKADLDAIDWKILRELQADGRMTNVELSNRVGISAPPCLRRV
                KRLEEAGIIRGYRALLNAPALGLDVVAFCLVGLHHQADAELRTFAERTRGWPIVRDAW
                MVSGESDFLLHCLASDLGAFQTFVIEELTSTPNVDTVRTALTIRRVKDEGLVSFERS"
     gene       3093. .3887
                /gene="mlr2557"
     CDS        3093. .3887
                /gene="mlr2557"
                /codon_start=1
                /transl_table=11
                /product="glucosyl transferase"
                /protein_id="BAB49659.1"
                /db_xref="GI:14023052"
                /translation="MTRFASPVSALIICVNEADMIGPCLESIDFCAEIIIVDSGSTDG
                TIECVKSYIAKGYPIRLLHNDWPGFPRQRQFSLDHATQPWCLSIDPDERVDDQLRQSI
                VAITQAADDKISGWYIRRRDWLKGYGYAHRWVLHNRLMRLFRREGATIDLTARVHEAF
                QVPGETGTVEDGVLLHRREISVEEDLARANAYSSLKAVTLVEKGKKPGLVRLVLSPLG
                NFLKFYLAKRYFLCGRHGFVYSMSVMIYSFATEAKLYEASERDDPV"
     gene       complement(3903. .4556)
                /gene="mll2559"
     CDS        complement(3903. .4556)
                /gene="mll2559"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB49660.1"
                /db_xref="GI:14023053"
                /translation="MADKTGTPHPLTEPGVWIERIGGRVFPPHLPALFLDRDGTINVD
                TDYPSDPAEIVLRPQMLPAIATANRAGIPVVVVTNQSGIARGYFGWSAFAAVNGRVLE
                LLREEGVFVDMVLACAYHEAGVGPLAIPDHPMRKPNPGMLVEAGKRLALDLQRSLIVG
                DKLADMQAGKRAGLAQGWLVDGEAAVQPGFAIRPLRDSSELGDLLAAIETLGRDNRS"
     gene       complement(4549. .5121)
                /gene="mll2561"
     CDS        complement(4549. .5121)
                /gene="mll2561"
                /codon_start=1
                /transl_table=11
                /product="probable phosphoheptose isomerase"
                /protein_id="BAB49661.1"
                /db_xref="GI:14023054"
                /translation="MSELNDYLVRSAAAITAMVERDLTGEMERAASAVVTALSSGKAL
                LIAGNGGSASDAIHIAGELVGRFLKERKAYNVIALPANAAVLTAWGNDYGFDTVFSRQ
                VEAHGTAGGVLLAISTSGNSPSILAAAEQARMMDMTVIGLTGDTGGKLKPLCDILLNV
                PSTSTPIIQQGHLCLYHHLCEVVEARLSNG"
     gene       complement(5129. .6619)
                /gene="mll2562"
     CDS        complement(5129. .6619)
                /gene="mll2562"
                /codon_start=1
                /transl_table=11
                /product="ADP-heptose synthase"
                /protein_id="BAB49662.1"
                /db_xref="GI:14023055"
                /translation="MIKHNPPSPEPLHRAIARFGGVTVLVVGDLILDRFVNGVIERIS
                PEAPIPVLHGRGETSAMGGAGNVVANIVSLGARAIPVSVIGTDTAGDSLVRMLAELGA
                ETAGLSQQRGRMTSSKSRFSALNQQVLRFDEEEIKPLDETERAGLIRHFRAALAGAEI
                VILSDYGKGILLDGVAAELIAICREAGKPVLVDPKGRDYARYAGATAITPNRKELGEA
                VGHAVFADDEIVAAARELISAHGFDFVVATRSEKGMSVVGPDEARHIATQAREVFDVS
                GAGDTVIATFALALASGADPVAAASIANAAGGVVVGKRGTARLTVEELTGALFRSHGP
                TAHKDAILDASSAARMVAAWKEEGLSVGFTNGCFDILHAGHVSLLHAARSQCDRLVLG
                LNSDASVRRLKGPGRPVNDQHDRACVLAALASVDAVVVFEEDTPLALIEALLPDILVK
                GADYTIDTVVGADVVQKAGGRVVLVDLVAGKSTTGTIGKLRAGSTS"
     gene       complement(6734. .7366)
                /gene="mll2564"
     CDS        complement(6734. .7366)
                /gene="mll2564"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB49663.1"
                /db_xref="GI:14023056"
                /translation="MPSTELLIAFFATTAIFAYIPGPAMLYAAAQTLARGRWSGLTAA
                LGIHLGGYVHVFAAAAGLSVLFHAVPPLYMAVKLIGALYLIWLGVSLFRKRVEGDVAL
                PAIERKSARRAFFESITVEVLNPKTAIFFMAFLPQFIDASAALPVWLQFVVLGTIVNL
                MFSSADIVCVFLAGLVISRLRRSSRAQRVMQRAGGAVLVGLGVHVALQKS"
     gene       7499. .8449
                /gene="mlr2565"
     CDS        7499. .8449
                /gene="mlr2565"
                /codon_start=1
                /transl_table=11
                /product="ADP-L-glycero-D-mannoheptose-6-epimerase"
                /protein_id="BAB49664.1"
                /db_xref="GI:14023057"
                /translation="MIIVTGGAGMIGSNIVAALNAEGHDDILVVDDLTDGHKIANLAD
                LSIADYLDKDDFLTRIEAGALGRIEAVFHQGACSTTTEWNGKFMMEVNFAYSKRLLHA
                CQAPRVPFLYASSASVYGGGSEFREEPALERPLNVYAYSKKLFDDYVRRTVFDTDHSQ
                VAGLRYFNVYGPREAHKGAMASVAFHLFNQVERGENPRLFGAYDGFGPGEQSRDFIHV
                GDVADVNLWLWKRGSSGIFNCGTGRAQPFRAIAETVIDTLGKGEIEFIPFPDHLKGSY
                QSFTQADMSRLRAAGYNGQFRTVETGVRDYVEWLKAQRSS"
     gene       8422. .9429
                /gene="mlr2566"
     CDS        8422. .9429
                /gene="mlr2566"
                /codon_start=1
                /transl_table=11
                /product="ADP-heptose-LPS heptosyltransferase"
                /protein_id="BAB49665.1"
                /db_xref="GI:14023058"
                /translation="MAESPTILVIGPRWVGDMVMAQCLFSALKELYPNAAIDVLAPAW
                AAPLVKRMPEIRQQIDLPLKPGALEFTIRRRFGRLLRGRYDMAYVLPGSWKSALIPFF
                ARIPRRFGHLREMRYGLLTDIVPLPDAVKRRTAQAYFSLAKGGSFRAPHLTVDAGNQA
                ALLDRFGLAPRKFAALMPGAEFGPAKRWPSESYAGLARKFMDKGLKVALFGSKNDRDV
                TAEITALAPGVVDLAGQTRLEDAIDLIAAARLAVSNDSGLMHVAAGVGTPIVAVYGST
                SPENTPPLAERRELVWLHLSCSPCHQKVCPLGHLNCLKTLEVGQVATAAERLLELPAT
                A"
     gene       9426. .10400
                /gene="mlr2567"
     CDS        9426. .10400
                /gene="mlr2567"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB49666.1"
                /db_xref="GI:14023059"
                /translation="MKVLIVKTSSMGDVIHTFPAVEDASLHRPDISFDWCVEEAFAGI
                VALHPAIAGIHTVAIRRWRKALFNGGTWREAAALRRTLRECRFDLVIDAQGLLKSALV
                ARQARAPIAGFDRSSAREPSATLFYDVTYGVPRNLHAIERTRRLFGLALGYQPDLSTL
                ASGIVPPPGALAGITGKTAFLLHGTSREDKKWPVEDWIGTARLLVERGMTPVTTWSNE
                REKSVAEAIARAVPSTVVVPKSPLADIAAILGRSTLVIGADTGLTHLASAFGLPTVAV
                FLATEPGLTGPRGPYASTLLAVPGGRLTPAEVVAEADSLLNLKLQAPA"
     gene       complement(10535. .11008)
                /gene="mll2568"
     CDS        complement(10535. .11008)
                /gene="mll2568"
                /codon_start=1
                /transl_table=11
                /product="transcriptional elongation factor"

alignment_scores:
            Quality:  106.50              Length:    171
              Ratio:  1.315                 Gaps:      9
 Percent Similarity:  47.368   Percent Identity: 23.392

alignment_block:
 US-09-674-779-2 x AP003000/rev   ..
```

```
Align seg 1/1  to reverse of: AP003000  from: 1  to: 348077

     72 MetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGl 88
        :::|||         |||   |||:::|||
 339623 CTGCACACGCATGAAAAGGCCGAGATCGTC.................TA 339592

     88 nSerArgLeuGlyAsnTyrLeuProPro...MetSerGlnLeuLeuThrT 104
         :::|||   |||:::|||||||||      :::   ::::::
 339591 CAAGCGCAACGGCCGCTACCTTCCCGAAGGTCTGCGCAAGATCAACATCA 339542

    104 hrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro 120
           |||:::|||:::         :::|||||   ::::::   |||
 339541 TCCTGCGCGACTGGCGCCGC......AACGAGCCGACCAAGATGGATCCG 339498

    121 GluHisLeu.........TrpGlyGlnIleValProThrLeuHisLeuTy 134
        :::   |||         |||                     :::   ||
 339497 CGCCTGCTGGACCTGGTCTGG.....................GAGGCCTA 339469

    134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
        |:::::      :::        :::             |||:::
 339468 TCGCGAAAGCGGCGCCACCGACTATATC............CAGGTCGTGT 339431

    151 erValTyrArgAsnProGluLeuAsn..............GlnCysAla 162
           ||||||:::|||      |||               :::
 339430 GCGGCTACCGTTCGCCGGCGACCAACTCGATGCTGCGCAGCCGCAGCCGC 339381

    163 GlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTr 179
        |||   |||      |||:::|||:::   ::::::|||:::|||:::::
 339380 GGTGTCGCCGAGAAGAGCCAGCATATGCTCGGCAAGGCGATGGACTTCTA 339331

    179 pValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnA 196
        ::::|||   :::   :::|||:::            |||
 339330 CATTCCCGGCGTGCCGCTGAAGAAGCTGCGCAATATCGGCCTG....... 339288

    196 rgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeu 212
                                   :::   |||:::   |||:::
 339287 ..........................AAGATGCAGGGCGGTGGCGTC 339267

    213 GlyLeuTyrAlaThrGlyAla.........IleHisLeuAspThrGlnGl 226
        |||   |||   |||::::::         :::|||:::|||:::   ::
 339266 GGCTACTATCCGACCTCCGGTTCGCCATTCGTCCATATGGATGTCGGCAA 339217

    226 yPheArgLysTrp 230
        :   |||   |||
 339216 TGTGCGCCACTGG 339204

seq_name: gb_pr:AC068614

seq_documentation_block:
LOCUS       AC068614              164706 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-512N1 from 2, complete sequence.
ACCESSION   AC068614
VERSION     AC068614.5  GI:15145605
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164706)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 164706)
  AUTHORS   Kang,K., Abbott,A., Boyer,E. and Dixon,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-512N1
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 164706)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 164706)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 164706)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 9, 2001 this sequence version replaced gi:14626397.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0512N01
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-102G8, 2000 bp overlap.
            Actual start of this clone is at base position 1 of RP11-512N1;
            actual end is at base position 164706 of RP11-512N1.

            Polymorphisms exist between RP11-512N1 and RP11-102G8.  Data from
            AC032014 was used to finish this clone AC068614.
FEATURES                Location/Qualifiers
     source             1. .164706
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
                        /clone="RP11-512N1"
                        /clone_lib="RPCI-11"
     repeat_region      1. .138
                        /rpt_family="MER1_type"
     repeat_region      145. .804
                        /rpt_family="L1"
     repeat_region      836. .1669
                        /rpt_family="L1"
     repeat_region      1685. .1848
```

```
                /rpt_family="MER1_type"
repeat_region   1868. .2044
                /rpt_family="MER1_type"
repeat_region   2260. .2411
                /rpt_family="CR1"
repeat_region   2485. .3182
                /rpt_family="L2"
repeat_region   3172. .3293
                /rpt_family="L2"
repeat_region   3298. .3488
                /rpt_family="L2"
repeat_region   3458. .3484
                /rpt_family="AT_rich"
repeat_region   3617. .3718
                /rpt_family="MER1_type"
repeat_region   3719. .3886
                /rpt_family="Alu"
repeat_region   3887. .3946
                /rpt_family="MER1_type"
repeat_region   3954. .4018
                /rpt_family="MIR"
repeat_region   5044. .5194
                /rpt_family="MIR"
repeat_region   5700. .5780
                /rpt_family="MER1_type"
repeat_region   6317. .6551
                /rpt_family="MIR"
repeat_region   6824. .6871
                /rpt_family="CT-rich"
repeat_region   7157. .7256
                /rpt_family="L2"
repeat_region   8291. .8715
                /rpt_family="MaLR"
repeat_region   8744. .8909
                /rpt_family="L2"
repeat_region   8933. .8971
                /rpt_family="MIR"
repeat_region   9722. .9786
                /rpt_family="L2"
repeat_region   10431. .10489
                /rpt_family="L2"
repeat_region   10686. .10897
                /rpt_family="MIR"
repeat_region   12174. .12338
                /rpt_family="L1"
repeat_region   12430. .12505
                /rpt_family="L2"
repeat_region   12688. .12963
                /rpt_family="Alu"
repeat_region   13850. .13945
                /rpt_family="MIR"
repeat_region   14692. .15044
                /rpt_family="L2"
repeat_region   15938. .16906
                /rpt_family="L1"
repeat_region   17595. .18092
                /rpt_family="AcHobo"
repeat_region   18081. .18507
                /rpt_family="AcHobo"
repeat_region   18374. .18426
                /rpt_family="AT_rich"
repeat_region   19734. .19816
                /rpt_family="MIR"
repeat_region   19971. .20124
                /rpt_family="MER103"
repeat_region   20810. .20877
                /rpt_family="(TGGA)n"
repeat_region   21098. .21164
                /rpt_family="L2"
repeat_region   21865. .22233
                /rpt_family="L2"
repeat_region   22328. .22439
                /rpt_family="MIR"
repeat_region   22440. .22664
                /rpt_family="MIR"
repeat_region   22665. .22747
                /rpt_family="MIR"
repeat_region   23939. .24091
                /rpt_family="MIR"
repeat_region   24668. .24701
                /rpt_family="(TTTA)n"
repeat_region   24783. .25084
                /rpt_family="Alu"
repeat_region   25063. .25084
                /rpt_family="AT_rich"
repeat_region   25255. .25297
                /rpt_family="L2"
repeat_region   26227. .27186
                /rpt_family="ERV1"
repeat_region   27187. .27284
                /rpt_family="ERV1"
repeat_region   27285. .27610
                /rpt_family="ERV1"
repeat_region   27848. .28178
                /rpt_family="L2"
repeat_region   29215. .29249
                /rpt_family="AT_rich"
repeat_region   30350. .30386
                /rpt_family="AT_rich"
repeat_region   32688. .32732
                /rpt_family="MIR"
repeat_region   33373. .33665
                /rpt_family="(TA)n"
repeat_region   35134. .35187

alignment_scores:
            Quality:  106.00              Length:    260
              Ratio:  0.855                 Gaps:     12
 Percent Similarity:  47.692   Percent Identity: 22.308

alignment_block:
 US-09-674-779-2 x AC068614/rev   ..

 Align seg 1/1  to reverse of: AC068614  from: 1  to: 164706

       8 PheIleThrThrLeuIleSerSerMetLeuValAla......CysSerAl 22
         |||:::::::   |||   |||:::   |||:::         ||||||
   95720 TTTTTCTCTCTTCTGGCTTCCCAACCTCTGATCCTCTCCCATTGCTCATA 95671

      22 aProIleProThrAsnProGlnValSerProIleLysThrProSerValL 39
         |||:::|||::::::|||                              :
   95670 CCCACTCCCAGTCACCCCC.........................TTTA 95648

      39 euIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspGlu 55
         ::   |||     :::|||  |||   |||:::         |||:::
   95647 TATACACACATACACAAATTGAGGACAGTCATGTTGCTTAGTCAGACAAA 95598

      56 SerValSer...........HisValGlyLeuGlnAlaHisPheGluTh 68
         :::   :::           |||::::::   ::::::|||
   95597 GAAGAAACATTTCGCTGAAGGCACACAGCTGAAAGAGGTCAT........ 95556

      68 rTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnA 85
         |||   :::     :::     ::::::     |||   |||:::
   95555 .TGGTCATCCAGATCCAATAGAGAAGAGGAGGTTTCTGTGCCATATAAGA 95507

      85 laTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeu 101
         :::           ||||||      :::   |||      :::
   95506 GGTTTGC.........CTGGGAATGAGAGTTTCCCCATGTTGCAGGACT 95467

     102 LeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLe 118
         :::           ||||||::::::|||   |||         :::
   95466 TTC...........TCCTGGGAAAGTTGCAGACACACCAGGATGAGTTG 95429

     118 uProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrG 135
           |||   |||                             |||   |||:::
   95428 GTCACCCTGCCATTCTAGGCAAGCTGGTGAACAGCTTTTATTCTTGTTTT 95379
```

```
    135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
          :::          |||   ||||||   |||:::   :::   |||
  95378 ATTCTTGCTATCTCAGATTTATCCTTGAGGCCTCAGGTTCA...AGGCAT 95332

    152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
        ::::::|||:::   |||   ::::::          ::::::::::
  95331 ACATTTAGAAGAGGAGAGTGTACCAGGCTTCTCTTGTGTGTCTGCTTACA 95282

    168 rLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluI 185
        :          |||:::   |||   :::   :::   :::|||:::|
  95281 AATCAAAGAAAAGAACAAGATTATTACAGTG...ATTTCCAACCTCCAAA 95235

    185 leLysSerGlnAlaLeuTyrGluLeuGlnAsn.........ArgLeuCys 198
        ||           |||      ||||||:::            ||||||
  95234 TTACTTTTTACAACCTCCCTCCACTCCAGCATCCACATGGAAGTTTGTGT 95185

    199 GlnTyrTrpLeuGlu.....HisGlyGluAsnGlnAsnPheGlyLeuGl 213
        ::::::::::::|||      |||
  95184 AAATGGTATGTGGAAATCACACATCCC..................... 95158

    213 yLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysT 230
          ||||||   :::   :::::::::|||||||:::         :::|
  95157 ....TATGCTCAGAGTCAGGTTCAGGTGGACACAGATGTGGGTGGCTCCT 95112

    230 rp.............................GlyAlaGlnPheSer 235
        ||                            |||::::::   :::
  95111 GGGAAGATAAAGAAAATATTTTCTGGTGGAGGAGGGGAGGTAGCAAAACA 95062

    236 GluThrAsnSerIleCysArgHisValLeu 245
        |||:::||||||:::   :::|||::::::
  95061 GAAAATAATTCTTTTGTCAAGCACATAATG 95032
```

seq_name: gb_htg:AC098619

seq_documentation_block:

```
LOCUS       AC098619              184172 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-125E15, *** SEQUENCING IN PROGRESS
            ***, 82 unordered pieces.
ACCESSION   AC098619
VERSION     AC098619.3  GI:17973396
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 184172)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184172)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17064347.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHVG
                Center clone name: CH230-125E15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 147975 bases at least Q40
                Consensus quality: 158603 bases at least Q30
                Consensus quality: 166972 bases at least Q20
                Estimated insert size: 151344; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 82 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5267: contig of 5267 bp in length
            *     5268     5367: gap of unknown length
            *     5368     9333: contig of 3966 bp in length
            *     9334     9433: gap of unknown length
            *     9434    14354: contig of 4921 bp in length
            *    14355    14454: gap of unknown length
            *    14455    18618: contig of 4164 bp in length
            *    18619    18718: gap of unknown length
            *    18719    23127: contig of 4409 bp in length
            *    23128    23227: gap of unknown length
            *    23228    25777: contig of 2550 bp in length
            *    25778    25877: gap of unknown length
            *    25878    28855: contig of 2978 bp in length
            *    28856    28955: gap of unknown length
            *    28956    32638: contig of 3683 bp in length
            *    32639    32738: gap of unknown length
            *    32739    37330: contig of 4592 bp in length
            *    37331    37430: gap of unknown length
            *    37431    40552: contig of 3122 bp in length
            *    40553    40652: gap of unknown length
            *    40653    43958: contig of 3306 bp in length
            *    43959    44058: gap of unknown length
```

```
*    44059    46235: contig of 2177 bp in length
*    46236    46335: gap of unknown length
*    46336    49232: contig of 2897 bp in length
*    49233    49332: gap of unknown length
*    49333    52701: contig of 3369 bp in length
*    52702    52801: gap of unknown length
*    52802    55361: contig of 2560 bp in length
*    55362    55461: gap of unknown length
*    55462    58603: contig of 3142 bp in length
*    58604    58703: gap of unknown length
*    58704    62092: contig of 3389 bp in length
*    62093    62192: gap of unknown length
*    62193    65054: contig of 2862 bp in length
*    65055    65154: gap of unknown length
*    65155    67142: contig of 1988 bp in length
*    67143    67242: gap of unknown length
*    67243    69439: contig of 2197 bp in length
*    69440    69539: gap of unknown length
*    69540    71427: contig of 1888 bp in length
*    71428    71527: gap of unknown length
*    71528    74782: contig of 3255 bp in length
*    74783    74882: gap of unknown length
*    74883    78104: contig of 3222 bp in length
*    78105    78204: gap of unknown length
*    78205    81915: contig of 3711 bp in length
*    81916    82015: gap of unknown length
*    82016    84805: contig of 2790 bp in length
*    84806    84905: gap of unknown length
*    84906    87400: contig of 2495 bp in length
*    87401    87500: gap of unknown length
*    87501    89572: contig of 2072 bp in length
*    89573    89672: gap of unknown length
*    89673    91932: contig of 2260 bp in length
*    91933    92032: gap of unknown length
*    92033    93201: contig of 1169 bp in length
*    93202    93301: gap of unknown length
*    93302    95058: contig of 1757 bp in length
*    95059    95158: gap of unknown length
*    95159    97729: contig of 2571 bp in length
*    97730    97829: gap of unknown length
*    97830   100033: contig of 2204 bp in length
*   100034   100133: gap of unknown length
*   100134   101773: contig of 1640 bp in length
*   101774   101873: gap of unknown length
*   101874   104074: contig of 2201 bp in length
*   104075   104174: gap of unknown length
*   104175   106043: contig of 1869 bp in length
*   106044   106143: gap of unknown length
*   106144   108722: contig of 2579 bp in length
*   108723   108822: gap of unknown length
*   108823   110209: contig of 1387 bp in length
*   110210   110309: gap of unknown length
*   110310   111720: contig of 1411 bp in length
*   111721   111820: gap of unknown length
*   111821   113557: contig of 1737 bp in length
*   113558   113657: gap of unknown length
*   113658   115613: contig of 1956 bp in length
*   115614   115713: gap of unknown length
*   115714   117216: contig of 1503 bp in length
*   117217   117316: gap of unknown length
*   117317   118478: contig of 1162 bp in length
*   118479   118578: gap of unknown length
*   118579   120669: contig of 2091 bp in length
*   120670   120769: gap of unknown length
*   120770   123348: contig of 2579 bp in length
*   123349   123448: gap of unknown length
*   123449   124599: contig of 1151 bp in length
*   124600   124699: gap of unknown length
*   124700   127087: contig of 2388 bp in length
*   127088   127187: gap of unknown length
*   127188   129104: contig of 1917 bp in length
*   129105   129204: gap of unknown length
*   129205   131124: contig of 1920 bp in length
*   131125   131224: gap of unknown length
*   131225   133021: contig of 1797 bp in length
*   133022   133121: gap of unknown length
*   133122   134279: contig of 1158 bp in length
*   134280   134379: gap of unknown length
*   134380   136332: contig of 1953 bp in length
*   136333   136432: gap of unknown length
*   136433   137878: contig of 1446 bp in length
*   137879   137978: gap of unknown length
*   137979   140375: contig of 2397 bp in length
*   140376   140475: gap of unknown length
*   140476   141842: contig of 1367 bp in length
*   141843   141942: gap of unknown length
*   141943   143652: contig of 1710 bp in length
*   143653   143752: gap of unknown length
*   143753   145714: contig of 1962 bp in length
*   145715   145814: gap of unknown length

alignment_scores:
           Quality:  105.50              Length:    241
             Ratio:  0.977                 Gaps:     13
 Percent Similarity:  44.813   Percent Identity: 21.162

alignment_block:
 US-09-674-779-2 x AC098619   ..

 Align seg 1/1  to: AC098619  from: 1  to: 184172

      5 AsnGlnTyrPheIleThrThrLeuIle..................... 13
        |||:::::::|||:::::::::|||:::
  36176 AATAAACATTTCCTCTCCAGTTTGCTTTAGGTCGGGGAGTTTATAGACTA 36225

     14 ....SerSerMetLeuValAlaCysSer.................AlaP 23
             ::::::::::::|||:::||||||                 ||||
  36226 CTAAACGGGCATTATTGTTGTATGTAGTTTTTGCACCCAGCGTTAAGCTC 36275

     23 roIleProThrAsnProGlnValSerProIleLysThrProSerValLeu 39
        ||:::   |||||||||:::   :::   :::::::::   |||
  36276 CAGTCACTACCAACCCCCATCCAGAATCCCTTGAATCCACAAGT...... 36319

     40 IleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluSe 56
                              :::||||||||||||   |||   :::::
  36320 .................AAAAATCACCACACACACACACACACACAC 36351

     56 rValSerHisValGlyLeuGlnAlaHis..................... 65
        :   :::|||:::     ::::::|||.
  36352 ACACACACACACACACACACACACACACACACACTTAATTTAAAATGCCTTT 36401

     66 .....PheGluThrTrp............LeuGlnMetHisHisAlaThr 76
             ::::::   |||           |||:::   |||:::|||:::
  36402 ATTAACTCAATGACTGGGCACTGCTTAAGCTCTCTCCCCACCAGGCAAAT 36451

     77 LysGlnGluVal...ValArgTyrGlnAlaTyrLeuGlnSerArgLeuGl 92
        :::::::::|||   :::   |||               :::   |||
  36452 TCGGATAGAGTGGCCATTGGCTACTGAATCCAAAAACTCACATGGCTGGT 36501

     92 yAsnTyrLeu.................................ProP 97
        :::::::::                                  ||||
  36502 TGGCTTCATCTCTTGCAGCTACTACTCTTTCTTCTCCCTCCTTGACCCCC 36551

     97 roMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHis 113
        ||                                 |||   |||
  36552 CT.............................CAATTCTGCCTGTGT 36568

    114 GluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProTh 130
           |||   ::::::||||||      |||   :::   ::::::
  36569 CCTCCTCTAAGTGTCCCACCCTGCCCTCTG...TCACTGGTTGCTTGAAT 36615

    130 rLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnT 147
        |||   :::   |||:::                        |||   :::
  36616 TTTATTAATGGATCAAGAA.....................CCAATTGGGG 36644

    147 hrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGly 163
```

```
          |||      |||||||
    36645 AACAGGCTCTTAGTGTC.............................. 36661

      164 GlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpVa 180
          ::::::|||||||||                 |||::::::
    36662 ...AGAACCATCTCTAAACAT.............TATGATGTGTATTG 36693

      180 lProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgL 197
          |||   |||   :::|||:::|||           :::::::::
    36694 TCCAGCACTCCCATTAAAAAAGCAA...........ATTAGAACACTGT 36731

      197 euCysGlnTyrTrpLeuGluHis 204
          :::|||||||||::::::
    36732 ATTTTAGCTATTGGCTACACTAT 36754

seq_name: gb_htg:AC020643

seq_documentation_block:
LOCUS       AC020643              162527 bp    DNA     linear   HTG 30-AUG-2001
DEFINITION  Homo sapiens chromosome 03 clone RP11-486C24, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC020643
VERSION     AC020643.4  GI:9887661
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162527)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162527)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Aug 24, 2000 this sequence version replaced gi:8389446.
            -------------- Genome Center
                           Center: Genome Therapeutics Corperation
                           Center code: GTC
                           Web site: http://www.genomecorp.com/
                           Contact: gtc-seqcenter@genomecorp.com
                       -------------- Project Information
                           Center project name: hg085
                       -------------- Summary Statistics
                           Sequencing vector: N/A
                           Chemistry: Dye-terminator Big Dye; 100% of reads
                           Assembly program: Phrap; version 990315
                           Consensus quality: 149308 bases at least Q40
                           Consensus quality: 156323 bases at least Q30
                           Consensus quality: 157579 bases at least Q20
                           Insert size: 161926; sum-of-contigs
                           Quality coverage: 5.6x in Q20 bases; sum-of-contigs
                       --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2231: contig of 2231 bp in length
            *     2232     2331: gap of unknown length
            *     2332     7785: contig of 5454 bp in length
            *     7786     7885: gap of unknown length
            *     7886    20723: contig of 12838 bp in length
            *    20724    20823: gap of unknown length
            *    20824    41799: contig of 20976 bp in length
            *    41800    41899: gap of unknown length
            *    41900    62173: contig of 20274 bp in length
            *    62174    62273: gap of unknown length
            *    62274    93394: contig of 31121 bp in length
            *    93395    93494: gap of unknown length
            *    93495   129282: contig of 35788 bp in length
            *   129283   129382: gap of unknown length
            *   129383   162527: contig of 33145 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="03"
                     /clone="RP11-486C24"
                     /clone_lib="RPCI-11"
     misc_feature    1. .2231
                     /note="assembly_name:Contig6"
     misc_feature    2332. .7785
                     /note="assembly_name:Contig7
                     clone_end:SP6"
     misc_feature    7886. .20723
                     /note="assembly_name:Contig8"
     misc_feature    20824. .41799
                     /note="assembly_name:Contig9"
     misc_feature    41900. .62173
                     /note="assembly_name:Contig10"
     misc_feature    62274. .93394
                     /note="assembly_name:Contig11"
     misc_feature    93495. .129282
                     /note="assembly_name:Contig12"
     misc_feature    129383. .162527
                     /note="assembly_name:Contig13
                     clone_end:T7"
BASE COUNT    49414 a  32608 c  33570 g  46221 t    714 others
ORIGIN


alignment_scores:
            Quality:  103.50              Length:    286
              Ratio:  0.755                 Gaps:     14
 Percent Similarity:  47.902   Percent Identity: 18.881

alignment_block:
 US-09-674-779-2 x AC020643/rev   ..

 Align seg 1/1  to reverse of: AC020643  from: 1  to: 162527

      15 SerMetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSe 31
         ::::::|||:::   :::::::|||   :::|||   :::         ||
   74680 ACTCTCCTGGCACAAGCTGAAGCCACATTCCCTCCATCCAGCCTCCCCAG 74631

      31 rProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspH 48
         ||||      :::   ||||||::::::|||
   74630 CCCCCATCTGTCTGTCTCTGTCTTCCTTACA..................... 74600

      48 isHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAla 64
           |||||||||   |||   ::::::   :::|||:::      :::|||
   74599 ..CACACACACACACACACACACACACACACACACACACACACACATGCT 74552

      65 .....................HisPheGluThrTrpLeu.........Gl 71
                              ::::::::::|||::::::         ||
   74551 TAGGAGAAACAGGAACCAGAGGAGATTCAGACATTTATTTTGCTTTGGCA 74502

      71 nMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuG 88
         |:::::|||      |||:::      :::   :::         |||:
   74501 AATCTACCACTGACCTAAGAAGCCAGGACTTTTGTTTGCT......TTAG 74458

      88 lnSerArgLeuGly.............AsnTyrLeuProProMetSer... 99
         ::|||::::::|||              :::|||||      :::|||
   74457 AATCTCAGATGGGAAAAGGCACCCTAAGTTATTTATTTGTTTTTTCATGA 74408

     100 ............GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGl 112
                     :::::::|||   ::::::::::::|||||||||||||
   74407 ATTCTTTTTTATTCTGTACTTAGAGTCAGTCAGGGATGGCAAGCATGTGG 74358
```

```
   112 y.....His......G 114
       |              |||     :
 74357 TGTCGGCCTGTAGGGCAGCTGTGTACATCTGGGAGACATTCATAGTGTGA 74308

   114 luProTyrGlnLeuProProGluHisLeuTrpGlyGlnIle..... 127
       ::      ::::::   |||:::|||      :::::::::
 74307 GCACTGACTCTGTCATTCCCCAGCATAGAAGGAGCAATGTAGCTGGTGCT 74258

   128 .....ValProThrLeuHisLeuTyrGlnAs 136
                        :::|||   :::         :::
 74257 TCTGTGCCCCTATGAGCAGCCCAAATTCCAAAAATTCCATGGGAAGAAGG 74208

   136 p......LeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
         |||   ::::::   :::::::|||:::::::::::   |||:
 74207 GACACGGCTGGCTGAGCACACTGTGATTCCCACCACAGCTGAGTGCAGGG 74158

   151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMet 167
       ::                                          ::::::
 74157 AT.....CCTGGGGTG 74147

   168 SerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGl 184
       ::::::::::::::|||   ::::::      :::      |||   ||
 74146 AAAAATGAGATTGTAAACATCACCCTGGCAGCATTTTCATCAGATGAAGA 74097

   184 uIleLysSer.....GlnAlaLeuTyrGluLeuGlnAsnArgL 197
       |:::::::::            ::::::   :::|||   |||::::::
 74096 GATGGAAGGCCTGCACGATCTCAAGGGAGCCTGGGAAAATCAACACAATG 74047

   197 euCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGly 213
         |||   |||:::|||:::   |||:::|||:::|||
 74046 CCTGCGTATATTTTCTAGAT...GGGCAAAATAAAAAC..... 74012

   214 LeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTr 230
                                                   |||
 74011 .....AAACC 74007

   230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuProL 247
             :::|||:::|||   ::::::      :::|||:::
 74006 AAAAAAGAGGTTTAAGGAAGACGATGATGCAAGAAAGCACATCAGCACAA 73957

   247 ysAsnLys 249
         ||||||
 73956 TAAACAAA 73949
```

seq_name: gb_htg:AC027289

seq_documentation_block:

LOCUS AC027289 173797 bp DNA linear HTG 06-SEP-2000

DEFINITION Homo sapiens chromosome 3 clone RP11-96A6, WORKING DRAFT SEQUENCE, 16 unordered pieces.

ACCESSION AC027289

VERSION AC027289.9 GI:9966564

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 173797)

AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 173797)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (30-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Sep 5, 2000 this sequence version replaced gi:9438443.

```
---------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
    Center project name: HAOY
    Center clone name: RP11-96A6
---------- Summary Statistics
    Sequencing vector: M13; L08821
    Chemistry: Dye-primer Bodipy: 55% of reads
    Chemistry: Dye-terminator Big Dye: 45% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 158546 bases at least Q40
    Consensus quality: 164386 bases at least Q30
    Consensus quality: 167376 bases at least Q20
    Estimated insert size: 169737; sum-of-contigs estimation
    Quality coverage: 0x in Q20 bases; agarose-fp estimation
    Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation
----------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*      1    34548: contig of 34548 bp in length
*  34549    34648: gap of unknown length
*  34649    55498: contig of 20850 bp in length
*  55499    55598: gap of unknown length
*  55599    75479: contig of 19881 bp in length
*  75480    75579: gap of unknown length
*  75580    95887: contig of 20308 bp in length
*  95888    95987: gap of unknown length
*  95988   109010: contig of 13023 bp in length
* 109011   109110: gap of unknown length
* 109111   118399: contig of 9289 bp in length
* 118400   118499: gap of unknown length
* 118500   128880: contig of 10381 bp in length
* 128881   128980: gap of unknown length
* 128981   138897: contig of 9917 bp in length
* 138898   138997: gap of unknown length
* 138998   145661: contig of 6664 bp in length
* 145662   145761: gap of unknown length
* 145762   152105: contig of 6344 bp in length
* 152106   152205: gap of unknown length
* 152206   157871: contig of 5666 bp in length
* 157872   157971: gap of unknown length
* 157972   161046: contig of 3075 bp in length
* 161047   161146: gap of unknown length
* 161147   164725: contig of 3579 bp in length
* 164726   164825: gap of unknown length
* 164826   169473: contig of 4648 bp in length
* 169474   169573: gap of unknown length
* 169574   171882: contig of 2309 bp in length
* 171883   171982: gap of unknown length
```

```
          *   171983  173797: contig of 1815 bp in length.
FEATURES             Location/Qualifiers
     source          1. .173797
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-96A6"
BASE COUNT     51157 a   35485 c   34596 g   50978 t    1581 others
ORIGIN


alignment_scores:
            Quality:  103.50              Length:    286
              Ratio:  0.755                Gaps:     14
 Percent Similarity:  47.902   Percent Identity: 18.881

alignment_block:
 US-09-674-779-2 x AC027289/rev   ..

 Align seg 1/1  to reverse of: AC027289  from: 1  to: 173797

      15 SerMetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSe 31
         ::::::|||:::   ::::::|||   :::|||   :::         ||
  155725 ACTCTCCTGGCACAAGCTGAAGCCACATTCCCTCCATCCAGCCTCCCCAG 155676

      31 rProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspH 48
         ||||      :::   ||||||::::::|||
  155675 CCCCCATCTGTCTGTCTCTGTCTTCCTTACA..................... 155645

      48 isHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAla 64
           |||||||||   |||   ::::::   :::|||:::      :::|||
  155644 ..CACACACACACACACACACACACACACACACACACACACACACATGCT 155597

      65 .....................HisPheGluThrTrpLeu.........Gl 71
                             :::::::::|||:::::::         ||
  155596 TAGGAGAAACAGGAACCAGAGGAGATTCAGACATTTATTTTGCTTTGGCA 155547

      71 nMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuG 88
         |:::::::|||      |||:::      :::   :::         |||:
  155546 AATCTACCACTGACCTAAGAAGCCAGGACTTTTGTTTGCT......TTAG 155503

      88 lnSerArgLeuGly.............AsnTyrLeuProProMetSer... 99
         ::|||::::::|||             :::||||||      :::|||
  155502 AATCTCAGATGGGAAAAGGCACCCTAAGTTATTTATTTGTTTTTTCATGA 155453

     100 ............GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGl 112
                    ::::::|||   ::::::::::::|||||||||||||||
  155452 ATTCTTTTTTATTCTGTACTTAGAGTCAGTCAGGGATGGCAAGCATGTGG 155403

     112 y.....................................His......G 114
         |                                     |||      :
  155402 TGTCGGCCTGTAGGGCAGCTGTGTACATCTGGGAGACATTCATAGTGTGA 155353

     114 luProTyrGlnLeuProProGluHisLeuTrpGlyGlnIle......... 127
         ::      ::::::   |||:::|||      :::::::::
  155352 GCACTGACTCTGTCATTCCCCAGCATAGAAGGAGCAATGTAGCTGGTGCT 155303

     128 .......................ValProThrLeuHisLeuTyrGlnAs 136
                               :::|||   :::         :::
  155302 TCTGTGCCCCTATGAGCAGCCCAAATTCCAAAAATTCCATGGGAAGAAGG 155253

     136 p......LeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
               |||   ::::::   ::::::|||:::::::::::   |||:
  155252 GACACGGCTGGCTGAGCACACTGTGATTCCCACCACAGCTGAGTGCAGGG 155203

     151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMet 167
         ::                                      ::::::
  155202 AT.....................................CCTGGGGTG 155192

     168 SerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGl 184
         ::::::::::::::|||   ::::::      :::      |||   ||
  155191 AAAAATGAGATTGTAAACATCACCCTGGCAGCATTTTCATCAGATGAAGA 155142

     184 uIleLysSer.............GlnAlaLeuTyrGluLeuGlnAsnArgL 197
         |:::::::::             ::::::   :::|||   |||::::::
  155141 GATGGAAGGCCTGCACGATCTCAAGGGAGCCTGGGAAAATCAACACAATG 155092

     197 euCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGly 213
           |||   |||:::|||:::   |||:::|||:::|||
  155091 CCTGCGTATATTTTCTAGAT...GGGCAAAATAAAAAC............. 155057

     214 LeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTr 230
                                                        |||
  155056 .........................................AAACC 155052

     230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuProL 247
               :::|||:::|||   ::::::      :::|||:::
  155051 AAAAAAGAGGTTTAAGGAAGACGATGATGCAAGAAAGCACATCAGCACAA 155002

     247 ysAsnLys 249
           ||||||
  155001 TAAACAAA 154994

seq_name: gb_ba:AE005826

seq_documentation_block:
LOCUS       AE005826               12528 bp    DNA     linear   BCT 28-MAR-2001
DEFINITION  Caulobacter crescentus  section 152 of 359 of the complete genome.
ACCESSION   AE005826 AE005673
VERSION     AE005826.1  GI:13422890
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 12528)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
            Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R.,
            Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
            Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
            Fraser,C.M.
  TITLE     Complete genome sequence of Caulobacter crescentus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
  MEDLINE   21173698
REFERENCE   2  (bases 1 to 12528)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
            Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
            Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
            Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
            Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12528
                     /organism="Caulobacter crescentus"
                     /db_xref="taxon:69394"
     gene            67. .717
                     /gene="CC1512"
     CDS             67. .717
                     /gene="CC1512"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK23491.1"
                     /db_xref="GI:13422891"
                     /translation="MHRRKLLQWGLGAIGAGLLPLAARADDDIIGALLEGKMPQATPV
                     PPPTTVAATVASIDPPALKPAVDPRWVHLHNVHTGEKLEAVYWENGDYVPDAVSALDK
                     VLRDYRNDEVHPIDRGLYDLLDQIARKTQSKGPFQVISGYRSPATNRLLSKRSGEVAK
```

```
                KSLHMDGKAMDIFLEDVELKHVRAAALDLSVGGVGYYPTSNFVHVDVGPVRKWTGT"
gene            complement(796. .1233)
                /gene="CC1513"
CDS             complement(796. .1233)
                /gene="CC1513"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK23492.1"
                /db_xref="GI:13422892"
                /translation="MSPAMCQQACAAAEAACRTLGGRSGLARPRGHARSRRVVGDLRI
                ELRRGSRLTDRRIGGDFRRGRDGLGLWRRRLCRLGGHRGRCRIIGGGAGLGGSVLVAA
                AAGRQEQNGPEGGQHADDETQQEQAKRIATTLAHLNGPPSTPT"
gene            1336. .1635
                /gene="CC1514"
CDS             1336. .1635
                /gene="CC1514"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK23493.1"
                /db_xref="GI:13422893"
                /translation="MVGVCHCSRCRKAGSSVYAYVRAEAFFWVAGRDLVARYAPTPPL
                RFNRCFCARCGTALGDPFSGRVLAIAASCLDDGVRLTPDFHEYVADSPSWRRPEA"
gene            complement(1705. .2745)
                /gene="CC1515"
CDS             complement(1705. .2745)
                /gene="CC1515"
                /note="identified by match to PFAM  protein family HMM
                PF01449"
                /codon_start=1
                /transl_table=11
                /product="phosphate ABC transporter, periplasmic
                phosphate-binding protein PstS, putative"
                /protein_id="AAK23494.1"
                /db_xref="GI:13422894"
                /translation="MNKLIGAVATVALLAVADQAHAARDQIWAVGSSTVFPFSTRVAE
                NFAKKTGKKSPRIEALGTGGGIKMFCGGAGEKFPDIANASRQMKKSEFLACQKAGVKD
                IVEIKIGFDGIVVASNKKGPDFNFKLEQLYLGLADQTLRGGQVVKQPYKSWSEVGPGL
                PNARILAYGPPPTSGTRDAWIELAIEGGAGKLPTLAKMKAADEKKFKATVSPMRTDGG
                WVDAGENDNAIVGTIEKTPNALGVFGYSFLEENGSRIKGASINGVKPTAQTIASGQYP
                LSRSLYIYVKKSQVGVTPGLKEFVSEFVSDAATGRGGYLQNRGMIPLPPAVHIQMKQK
                ANALTPMPAPKN"
gene            complement(2870. .5254)
                /gene="CC1516"
CDS             complement(2870. .5254)
                /gene="CC1516"
                /note="identified by match to  protein family HMM"
                /codon_start=1
                /transl_table=11
                /product="penicillin-binding protein, 1A family"
                /protein_id="AAK23495.1"
                /db_xref="GI:13422895"
                /translation="MSDPTDPADDVPAAPAPKPRRPRKPRAQTVMTEAAGLDPAAPED
                AAQKPRRPRKSRRAQPEAVAAPEAQTPTTSTAATDDADADLFEPLPSPEPDAEPVSVE
                ADPAETLEADAAADLESPTEPEVEPAQDIPPEAPAEPPTTAEAEDAPEAKAETSVPTL
                DEPEKVAAKPPIWKRPAWLIGAGAAVLLVIVLIASLFWSLPLSRALEPLDNSAIVLVA
                EDGSPIARRGSYKEAPIDVAKLPPYVPGAFIAIEDRRFYSHMGVDPKAIARAVGRNAQ
                AGGVSEGGSTITQQLAKNAFLSSDRNLRRKAQEALIAVYLEARLSKDEILSRYLSSVY
                FGDGAFGLRAAARHYFGKPPEQLSIGEAAMLAGLVKAPSRLAPTNNIEGARERMRVVL
                GAMVETKTITQAEADAVGEILPVEAEEKLPTGSYFADWALPQARALIGARYGETIVKT
                TLDPELQEKAERILNDYIERDGQVLNVTQGALVAMRRDGRVVAMVGGRDYKQSQFNRA
                DAERQPGSAFKLFVYLAALREGMTVTTPILDTPVQVSGYMPKNHEGKYHAREVPLITA
                FAGSSNVAAVRLAHDLGPAKVIKVARDLGVTEEIPSDLTMALGTGPMSLTRLTAAYAS
                VAAGEYPIVPHGVTEFKKPKTKAYDPKVLANMRDLLRSATHRGTGVEAAIPGAFGKTG
                TTQDYHDAIFVGYVEDLVVGVWLGNDDNSSMNGVVGGGEPARIWKAFMLSALGRDVAP
                VVEEDPLEDLGLPLEGELGPDGLPLAPFTSPTDEPAPPANSGVPLAPPPPEPRP"
gene            complement(5540. .7621)
                /gene="CC1517"
CDS             complement(5540. .7621)
                /gene="CC1517"
                /note="identified by match to  protein family HMM"
                /codon_start=1
                /transl_table=11
                /product="TonB-dependent receptor"
                /protein_id="AAK23496.1"
                /db_xref="GI:13422896"
                /translation="MTSKALLFALCAGVSQFAFADLAVAQDKTDEVEGVVVTGSRAAS
                ASINGVVIDPMRLPQSVRVLDEALIVDTGVTNLSDLFDFAGGMARQNSFGGAWDAYAI
                RGFSGDINQGPDLLVNRFTANRGFNARRDVATVERFQVLKGPASALSGKGEPGGSINI
                VTKAPTETAQGSGELSYGSFDAKRIMGDLSGPLGGGVSARMIAVYQDTDGWRDHVGSD
                RLLLAPSLAWTPSDDLRLLYQLEANTVHFVHDRGLVAVAGNGKALPRERFLGEPNDGD
                ITQKTLQHQLTTTYNFSPSVAVEAGVQYRDGSFRGQSTHNGALIGTQLRRQLRIHDYT
                WDDLSGRIEVSFDGKLGGLEHQLRAGADAFTYEQHRIFYRFNPTAATPYAIDILNPVY
                GQAKPVAPLNQNVLEELRGESLYLQDLVTLNSQFTLLVGVRQDWIRQTNTNYRNNTVT
                RQSPSQASPRAALTWAPNEAFSAYVSWGRSFRYNQGSDAVGGAFPPEKGEAWEAGVKW
                DLAGRLTGTASLFRIDKENILVNDPANSGFFIPVGAARSQGFEAETNLRLPKGITATA
                VYAYTDTEITRDTRTNMIGSSLSNVPKHSGAVYANWRSDGDAPGSVTLGGGVVYVGER
                AGDDVNTGFKLPDYVTVRANLAYNVSKAVSLHLDVENLFDTYYLESSYSNVWITPGAP
                RTITGRLRVSF"
gene            7769. .8494
                /gene="CC1518"
CDS             7769. .8494
                /gene="CC1518"
                /note="identified by match to  protein family HMM"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter, ATP-binding protein"
                /protein_id="AAK23497.1"
                /db_xref="GI:13422897"
                /translation="MDGDKAIEIADLVVRYGDRAVIDHLALTVRAGEVYGLLGGNGAG
                KSTTLQAILGFVRPSAGRVRVVGRDPASEPAEVRRAIAYIPENVALYEHLTARENIDY
                FLRLAGLRPVADEINKAFGTVGLAEEAWDRRVSGFSKGMRQKTAIALAVLRRASVLLL
                DEPTSGLDPGATADFNALVSRLRGEGAAILIVTHDLIGVADIADRIGFLERGRISDEM
                VAEGPERFDLRALHNRFSRARAA"
gene            8491. .9939
                /gene="CC1519"
CDS             8491. .9939
                /gene="CC1519"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK23498.1"
                /db_xref="GI:13422898"
                /translation="MSVILSVTRNELRHLVRSRVALTGLILVALLTFAAALTSNAYQA
                DQAGLRERLQHAADQDFAAQPNRHPHRVVHFGHFAIRPASGLAAMDPGVEAYTGNMIF
                LEGHRQNSANFGDARQSSLLVRFGQLSPALVLQVIVPLLLVFIGAGMIAGERERGTLR
                QSLLAGISGGALLSGKALALGLAALAISIPAFGLLLWLGASGGARLSAVGLTALGYVA
                YLAFWTLLIIAVSATANRSRSALIGLVGAWAFMVVLVPRVAPEAAAALAPLPTRVETD
                IAVQRDLRSMGDSHNPNDPYFAAFKQKTLKQYGVTRVEDLPVNYRGLLAIEGEKLTSS
                LFDRYADQAFGAMEAQGRVMDALAVVSPTVAIRRVSMTLAETDLFAYRRFLEQAEAYR
                YDLVQRLNRLQATAVTAADDAAKSKNAAAEARSRISADHWQAMPHFDPKRPSAADVVG
                RAVPALVALLLWLAAGAVFAARAGHRLNRSAA"
gene            9936. .11291
                /gene="CC1520"
CDS             9936. .11291
                /gene="CC1520"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK23499.1"
                /db_xref="GI:13422899"
                /translation="MNALLREARFLLGLKSAVAAMVALFALTAVAVSAGLVEVARQEA
                AIARIGPQQASDVAAVTDWVSQGKDPGNAAYYTFHATSDPPSDLAFAAIGTRDVSPYV
                LRVRALGLEAQLYEGEIGNPEAALPGRFDFAFVLVYLAPLFVILLFHDLRSGEREAGR
                LRALEASVGAPSALWGARLTVRALGLFLALALPFFVGAFVAGTALWKVGAVLAVVLAY
                LAFWIVLCMVVARAARSSLANAMTLAAAWMTLTLIVPAIGHVAINSAIPLRQGMELGQ
                TQRTAVHRAWDIPKSESMDAFFRTHPQWRNTAPLGEAFHWKWYFAFHQVGDELAGDLS
                RQYRQGVLERAAWSERLGLVTPGVATQLALHRLAATDPAAQIAYQDRIRAFHGELRRF
                YYPYLFEDRPFGRADFARAPEYRAAPAAGSWPVALLGGVVLMTALVGGIGFLRGRRVA
                S"
gene            complement(11296. .12432)
```

```
                /gene="CC1521"
     CDS        complement(11296. .12432)
                /gene="CC1521"
                /note="similar to GB:D30759 PID:551631; identified by

alignment_scores:
            Quality:  103.00              Length:    237
              Ratio:  0.963                 Gaps:     12
 Percent Similarity:  45.148   Percent Identity: 23.629

alignment_block:
 US-09-674-779-2 x AE005826   ..

 Align seg 1/1  to: AE005826  from: 1  to: 12528

      28 ProGlnValSerProIleLysThrProSerValLeuIleThr..... 41
         ||||||::::::|||::      |||:::::::::   :::,
     181 CCTCAGGCCACGCCGGTTCCGCCGCCGACCACCGTCGCTGCGACGGTCGC 230

      42 .....LysAspLysIleGlyAspHisHisThrH 51
                      |||      :::      :::   :::|
     231 TTCGATCGATCCGCCCGCGCTGAAGCCTGCTGTGGATCCGCGCTGGGTTC 280

      51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
         ||   |||:::
     281 ATCTCCACAAC..................................... 291

      68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
                 :::|||   :::   |||   |||:::|||   :::::
     292 ............GTCCATACCGGCGAGAAGCTGGAAGCGGTCTACTGGGA 329

      84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuPro...ProMetSerG 100
         :                 |||:::|||:::|||      :::|||
     330 GAAC..............GGCGACTATGTTCCGGACGCCGTGAGTG 361

     100 lnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyr 116
           |||      :::   |||:::::::::           :::|||
     362 CGCTGGACAAGGTGCTGCGCGACTATCGCAACGACGAGGTCCATCCCATC 411

     117 GlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLe 133
         :::              |||      ||||||:::
     412 GATCGGGGCCTCTATGACCTGCTCGACCAGATCGCG.............. 447

     133 uTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleA 150
           :::      :::|||:::|||   :::           |||:::
     448 ....CGCAAGACCCAGAGCAAGGGGCCGTTC...........CAGGTGA 481

     150 rgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGly...... 164
           |||   ||||||:::|||      |||:::      :::
     482 TCTCGGGCTACCGCTCTCCGGCGACCAACCGTCTGCTGTCCAAGCGCAGC 531

     165 .........AlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIl 178
                 |||      |||   |||:::   ::::::|||:::||||||
     532 GGCGAGGTGGCCAAGAAGAGCCTGCACATGGATGGCAAGGCCATGGACAT 581

     178 eTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnA 195
         |:::::   |||:::|||:::|||                  |||:::
     582 CTTCCTGGAGGATGTCGAGCTCAAGCACGTCCGCGCTGCGGCTTTGGAT. 630

     195 snArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGly 211
                       |||:::   |||             |||
     631 ...................CTGAGCGTTGGC.............GGC 645

     212 LeuGlyLeuTyrAlaThr...GlyAlaIleHisLeuAspThrGlnGlyPh 227
         :::|||   |||   |||   :::   :::|||:::|||:::
     646 GTCGGCTACTATCCGACCAGCAACTTCGTCCATGTCGATGTCGGACCGGT 695

     227 eArgLysTrp...........GlyAlaGln.PheSerGluThrAsnSer 239
          |||||||||          ||||||       |||:::   ||||||
     696 CCGCAAGTGGACGGGGACCTGAGGCGCTACTCGCCTCTCAGCCGAATTCG 745

     240 IleCysArg 242
         |||   :::
```

```
     746 ATAACCAAG 754

seq_name: gb_htg:AC003656_4

seq_documentation_block:
WPCOMMENT
  Sequence split into 8 fragments  LOCUS AC003656 Accession AC003656
     Fragment Name      Begin     End
     AC003656_0             1     110000
     AC003656_1        100001     210000
     AC003656_2        200001     310000
     AC003656_3        300001     410000
     AC003656_4        400001     510000
     AC003656_5        500001     610000
     AC003656_6        600001     710000
     AC003656_7        700001     753000
Continuation (5 of 8) of AC003656 from base 400001 (AC003656 Homo sapiens clone P1 C1


alignment_scores:
            Quality:  102.00              Length:    326
              Ratio:  0.823                 Gaps:     20
 Percent Similarity:  38.037   Percent Identity: 23.926

alignment_block:
 US-09-674-779-2 x AC003656_4/rev   ..

 Align seg 1/1  to reverse of: AC003656_4  from: 1  to: 110000

      16 MetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSerPr 32
         ||||||:::|||||||:::::          |||      ::::::|||
   88874 ATGCTGCTAGCGTGCACTGGCACGCATTCTACGTGTACCGAAACCAGCAG 88825

      32 oIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisH 49
          :::                                          |||
   88824 GCTA.........................................CACT 88817

      49 isThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHis 65
               :::|||   :::::::::|||:::   |||        :::
   88816 CCTGCTCTGATCACAAACACACTGCCTCGCGGAGGGGG.........CGG 88776

      66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValAr 82
          :::   |||||||||:::::::::
   88775 ACACAATGCTGGCTGCAGCTGAGACGCCCACCG.............. 88743

      82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetS 99
          |||||||||          |||||||||:::::|||         :
   88742 .TACCAGGCCTTG......CCGAGACTGGGATCCTGGCTGTCCCACCCCA 88700

      99 erGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu... 114
         ::          :::::::::   |||         ||||||::::::
   88699 AGCCCACAGGGTCGGTGGGCCTGTCTGTGACACTATGTGGGAGGAGGGCG 88650

     115 ProTyrGlnLeuPro.....................ProGluHisLe 123
         |||      ||||||                          |||
   88649 CCAGGCTGGCTACCCAGTCTAGGGTGGGACCACAGCACTGAGGAGAAGCC 88600

     123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
          ||||||         |||         |||   |||   |||:::||||:
   88599 CTGGGGA......TGCCCACTGAGCCTCCTGGTACAGCCCCTCAGAAGCC 88556

     140 rgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyr......... 153
         ::         ||||||||||||||||      :::   :::
   88555 AC......TGCCCAGCTAACACACAGCCACCCGGTTTCTGGGGCGGAGCC 88512

     154 ........................ArgAsnProGluLeuAsnGln.. 160
                                  |||:::||||||      |||
   88511 AGGCCTGGGGCCAGTGCCCTCCTGAGCCGGCACCCAGAGGGCATGCAAAG 88462

     160 ................................................ 160

   88461 AGGCCCCAGGGCTGGGCCCCAGGAGAAACACTGACCCTCACAGCCCAGAG 88412
```

```
   161 ........CysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSer 174
               |||:::|||||||||||||:::|||         :::::::::
 88411 GAGGCCTTTGTGTGGGAGGAGCTGCCCTGAGCTGCCTTCGAGCCTCGAAA 88362

   175 AlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeu.. 190
       ::::::   :::|||                ::::::   ||||||
 88361 GGTGTCACCCTCTGG.............GGCGTCAGACCACAGGCAGCCCC 88324

   191 .....................TyrGluLeuGlnAsnArgLeuCysGln. 199
                            :::   |||||||:::      |||
 88323 AGGTGTGCTAAATGTGCTTCCCTGGCCCCTCCAGCATCCTCCCTGTGGTC 88274

   200 ....................................TyrTrpLeuGlu 203
                                          |||
 88273 TGTCCTGCCACGCCTGCCTGCTGGTTCCGGCCGGGTGACAGTGGGCACCT 88224

   204 ........................HisGlyGluAsn........... 207
                               ||||||||||:::
 88223 GCTGAGCGGGTCTCTGCAGCCACCCTGCACGGCGAACGGCCCGGCCAGTT 88174

   208 ........................GlnAsnPheGlyLeuGlyLeuT 215
                               ||||||   |||||||||   :
 88173 GAGGTCTCCAGTGGCCTCGCTGTGACCTCAGAACCAGGGACTGGGT...C 88127

   215 yrAlaThrGly.AlaIleHisLeuAspThrGlnGlyPheArgLysTrpGl 231
       ::   :::||| |||              |||||||||              ||
 88126 ACAGGTCAGGTTGCACCTCCTCGACCCACACAAGGC.............GG 88089

   231 y.....................AlaGlnPheSerGluThrAsnSerIleC 241
       |                     ||||||   |||      :::|||   |
 88088 CCACCTGCCCCAAGCCAAGTGTGCCCAGACCTCCGTCCTGAGCTCAGCCT 88039

   241 ys.........ArgHisValLeuPro 246
       ||         |||   :::||||||
 88038 GCAGCAGATGGCGCTCTCTGCTGCCC 88013

seq_name: gb_pr:AP001053

seq_documentation_block:
LOCUS       AP001053              161920 bp    DNA     linear   PRI 14-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21, clone:KB836E9, MX1-D21S171
            region, complete sequence.
ACCESSION   AP001053
VERSION     AP001053.1  GI:6693603
KEYWORDS    HTG.
SOURCE      Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:Keio
            BAC library clone:KB836E9.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161920)
  AUTHORS   Shimizu,N., Kudoh,J. and Shibuya,K.
  TITLE     Homo sapiens genomic DNA, chromosome 21, clone KB836E9, MX1-D21S171
            region
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 161920)
  AUTHORS   Shimizu,N., Kudoh,J. and Shibuya,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2000) to the DDBJ/EMBL/GenBank databases.
            Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
            Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
            (E-mail:nshimizu@dmb.med.keio.ac.jp, Tel:81-3-3351-2370,
            Fax:81-3-3351-2370)
FEATURES             Location/Qualifiers
     source          1. .161920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="FLEB14-14"
                     /cell_type="pre-pro-B cell"
                     /chromosome="21"
                     /clone="KB836E9"
                     /clone_lib="Keio BAC library"
BASE COUNT    35486 a  41730 c  42915 g  41789 t

ORIGIN

alignment_scores:
            Quality:  102.00              Length:    326
              Ratio:  0.823                Gaps:     20
 Percent Similarity:  38.037   Percent Identity: 23.926

alignment_block:
 US-09-674-779-2 x AP001053/rev   ..

 Align seg 1/1  to reverse of: AP001053  from: 1  to: 161920

     16 MetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSerPr 32
        ||||||:::|||||:::::         |||      ::::::|||
  59556 ATGCTGCTAGCGTGCACTGGCACGCATTCTACGTGTACCGAAACCAGCAG 59507

     32 oIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisH 49
        :::                                              |||
  59506 GCTA.....................................CACT 59499

     49 isThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHis 65
              :::|||  :::::::::|||:::  |||         :::
  59498 CCTGCTCTGATCACAAACACACTGCCTCGCGGAGGGGG.....CGG 59458

     66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValAr 82
        :::  |||||||||:::::::::
  59457 ACACAATGCTGGCTGCAGCTGAGACGCCCACCG..... 59425

     82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetS 99
        |||||||||           ||||||||||:::::|||          :
  59424 .TACCAGGCCTTG......CCGAGACTGGGATCCTGGCTGTCCCACCCCA 59382

     99 erGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu... 114
        ::         :::::::::   |||         ||||||::::::
  59381 AGCCCACAGGGTCGGTGGGCCTGTCTGTGACACTATGTGGGAGGAGGGCG 59332

    115 ProTyrGlnLeuPro.....................ProGluHisLe 123
        |||    ||||||                        |||
  59331 CCAGGCTGGCTACCCAGTCTAGGGTGGGACCACAGCACTGAGGAGAAGCC 59282

    123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
        ||||||      |||         |||   |||   |||:::||||:
  59281 CTGGGGA......TGCCCACTGAGCCTCCTGGTACAGCCCCTCAGAAGCC 59238

    140 rgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyr..... 153
        ::        ||||||||||||||||      :::   :::
  59237 AC......TGCCCAGCTAACACACAGCCACCCGGTTTCTGGGGCGGAGCC 59194

    154 .......................ArgAsnProGluLeuAsnGln.. 160
                                |||:::||||||      |||
  59193 AGGCCTGGGGCCAGTGCCCTCCTGAGCCGGCACCCAGAGGGCATGCAAAG 59144

    160 ..................................................... 160

  59143 AGGCCCCAGGGCTGGGCCCCAGGAGAAACACTGACCCTCACAGCCCAGAG 59094

    161 ........CysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSer 174
                |||:::|||||||||||||:::|||         :::::::::
  59093 GAGGCCTTTGTGTGGGAGGAGCTGCCCTGAGCTGCCTTCGAGCCTCGAAA 59044

    175 AlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeu.. 190
        ::::::   :::|||                ::::::   ||||||
  59043 GGTGTCACCCTCTGG.............GGCGTCAGACCACAGGCAGCCCC 59006

    191 .....................TyrGluLeuGlnAsnArgLeuCysGln. 199
                             :::   |||||||:::      |||
  59005 AAGTGTGCTAAATGTGCTTCCCTGGCCCCTCCAGCATCCTCCCTGTGGTC 58956

    200 ....................................TyrTrpLeuGlu 203
                                           |||
  58955 TGTCCTGCCACGCCTGCCTGCTGGTTCCGGCCGGGTGACAGTGGGCACCT 58906
```

```
     204 .....................HisGlyGluAsn..... 207
                              |||||||||:::
   58905 GCTGAGCGGGTCTCTGCAGCCACCCTGCACGGCGAACGGCCCGGCCAGTT 58856

     208 ......................GlnAsnPheGlyLeuGlyLeuT 215
                               ||||||   |||||||||   :
   58855 GAGGTCTCCAGTGGCCTCGCTGTGACCTCAGAACCAGGGACTGGGT...C 58809

     215 yrAlaThrGly.AlaIleHisLeuAspThrGlnGlyPheArgLysTrpGl 231
         ::   :::||| |||           |||||||||            ||
   58808 ACAGGTCAGGTTGCACCTCCTCGACCCACACAAGGC.....GG 58771

     231 y.....................AlaGlnPheSerGluThrAsnSerIleC 241
         |                    ||||||   |||      :::|||   |
   58770 CCACCTGCCCCAAGCCAAGTGTGCCCAGACCTCCGTCCTGAGCTCAGCCT 58721

     241 ys.........ArgHisValLeuPro 246
         ||        |||   :::||||||
   58720 GCAGCAGATGGCGCTCTCTGCTGCCC 58695

seq_name: gb_htg:AC079571

seq_documentation_block:
LOCUS       AC079571              245130 bp    DNA     linear   HTG 02-SEP-2000
DEFINITION  Mus musculus clone RP23-72P2, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC079571
VERSION     AC079571.1  GI:9964936
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 245130)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245130)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1764882
            Center clone name: RPCI-23_72P2
            -------
            Summary Statistics
            Consensus quality: 233470 bases at least Q40
            Consensus quality: 238342 bases at least Q30
            Consensus quality: 239406 bases at least Q20
            Estimated insert size: 224000; agarose-fp estimation
            Estimated insert size: 243730; sum-of-contigs estimation
            Quality coverage:  9.68 in Q20 bases; agarose-fp estimation
            Quality coverage:  8.9 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1669: contig of 1669 bp in length
            *     1670     1769: gap of unknown length
            *     1770     3529: contig of 1760 bp in length
            *     3530     3629: gap of unknown length
            *     3630     5387: contig of 1758 bp in length
            *     5388     5487: gap of unknown length
            *     5488    11251: contig of 5764 bp in length
            *    11252    11351: gap of unknown length
            *    11352    19136: contig of 7785 bp in length
            *    19137    19236: gap of unknown length
            *    19237    28159: contig of 8923 bp in length
            *    28160    28259: gap of unknown length
            *    28260    38716: contig of 10457 bp in length
            *    38717    38816: gap of unknown length
            *    38817    58311: contig of 19495 bp in length
            *    58312    58411: gap of unknown length
            *    58412    71497: contig of 13086 bp in length
            *    71498    71597: gap of unknown length
            *    71598    88819: contig of 17222 bp in length
            *    88820    88919: gap of unknown length
            *    88920   111636: contig of 22717 bp in length
            *   111637   111736: gap of unknown length
            *   111737   131908: contig of 20172 bp in length
            *   131909   132008: gap of unknown length
            *   132009   160558: contig of 28550 bp in length
            *   160559   160658: gap of unknown length
            *   160659   186795: contig of 26137 bp in length
            *   186796   186895: gap of unknown length
            *   186896   245130: contig of 58235 bp in length.
FEATURES             Location/Qualifiers
     source          1. .245130
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-72P2"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    65628 a  57950 c  55473 g  64679 t   1400 others
ORIGIN


alignment_scores:
            Quality:  102.00              Length:    203
              Ratio:  0.971                Gaps:     13
 Percent Similarity:  51.724   Percent Identity: 24.138

alignment_block:
 US-09-674-779-2 x AC079571/rev   ..

 Align seg 1/1  to reverse of: AC079571  from: 1  to: 245130

      49 HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHi 65
         |||||||||   |||  ::::::                   :::::::||
   56186 CATACACACACACACACACACA.................CACACACA 56155

      65 sPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValA 82
         |   :::   |||        :::::::::      :::::::|||:::
   56154 CACAAAATGTTGGCCCACTTGCTATAACTGC......AATAATGTCATAA 56111

      82 rgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMet 98
            :::      |||   |||::::::      :::||||||
   56110 CTCAGGAGCTTCCTCTGTAATCTCATATTAAAATACACCTACCACAAAGG 56061

      99 Ser...GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
         :::   ::::::::::::   :::|||:::|||:::|||
   56060 GAGCCAAAAGTGGTGGCCCTGTCAAGGCAATGGGAAGCATGG........ 56019

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrL 131
         |||:::::::|||                              :::   :::
   56018 ....TATGACTTTCCAGAG.....................CTACAGTCCC 55994

     131 euHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThr 147
            |||           :::::::|||      |||      |||::::::
   55993 CTCATAGGCAGGCTCATGTAAGATCA...GTGATATGCATGGCCTCAGCT 55947

     148 GlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGl 164
         |||         |||      :::   :::   :::::::::   :::::
   55946 CAG......TGTGTGGTGTCCAGTGAAAGCTGGCAAAGAGCACCATCAGC 55903

     164 yAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValP 181
         :         |||||||||   |||:::                |||   |
```

```
  55902 CCACAAAGAATCCAAGCATGGGACTCAC.............TGGGGTC 55868

    181 roAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeu 197
        ||   |||:::::::::::::|||
  55867 CAAAGCTCCAGCTGCAAGCTCAGCTC.....................TGG 55839

    198 CysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLe 214
        |||:::::::||||||:::|||:::   |||   ||||||   |||
  55838 TGCCATCATTGGTTACAGCACTCCCCAAACCCCAACTTTGTACTC...TG 55792

    214 uTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg.LysTrp 230
         :::         |||   |||   |||      :::   ||| |||
  55791 TCACAGACCATAAGCC...CACACTGATCTTACAGCAAGCAGAGAAGAGT 55745

    231 GlyAlaGlnPheSer..............GluThrAsnSerIleCysAr 242
        |||::::::::::|||             :::   :::|||:::|||::
  55744 GGTACAGACATGTCTGTGGACGAGTGAAGACAACAAGAAAGTGTCTGTGA 55695

    242 gHisVal 244
        :||||||
  55694 GCATGTA 55688
```

seq_name: gb_pr:AP001752

seq_documentation_block:

LOCUS AP001752 340000 bp DNA linear PRI 30-MAY-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 96/105.
ACCESSION AP001752 AL163297 BA000005
VERSION AP001752.1 GI:7768757
KEYWORDS .
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
TITLE The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium
JOURNAL Nature 405 (6784), 311-319 (2000)
MEDLINE 20289799
REFERENCE 2 (bases 1 to 340000)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
TITLE Direct Submission
JOURNAL Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
COMMENT On May 30, 2000 this sequence version replaced gi:7717416.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: hattori@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,
* e.mail: nshimizu@dmb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163297: Submitted (10-Apr-2000).

FEATURES Location/Qualifiers

```
     source          1. .340000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1. .39630
                     /note="Accession No. AP001051"
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="KB161B12, 5' partial"
                     /clone_lib="Keio BAC library"
     repeat_region   924. .2304
                     /note="L1M1"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   2307. .2592
                     /note="AluJo"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   2593. .2616
                     /note="(CAAAA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   2617. .3063
                     /note="L1PA16"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   3712. .3856
                     /note="L1MC/D"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   3927. .4224
                     /note="AluJb"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   4285. .4400
                     /note="L1MC/D"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   4657. .4874
                     /note="AluJb"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   4968. .5280
                     /note="L1ME"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   5798. .5987
                     /note="L1MD"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   5989. .6291
```

```
                /note="AluY"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
repeat_region   6302. .6441
                /note="FLAM_A"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
repeat_region   6530. .7150
                /note="MLT1E2"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
mRNA            complement(join(<7441. .7566,23941. .24073,26891. .27040,
                37908. .38011,50206. .50356,51680. .51939,53005. .53112))
                /gene="HSF2BP"
                /note="Accession No. AB007131"
                /product="heat shock transcription factor 2 binding
                protein"
gene            complement(7441. .53112)
                /gene="HSF2BP"
CDS             complement(join(<7441. .7566,23941. .24073,26891. .27040,
                37908. .38011,50206. .50356,51680. .51715))
                /gene="HSF2BP"
                /note="Accession No. AB007131
                partial"
                /codon_start=1
                /product="heat shock transcription factor 2 binding
                protein"
                /protein_id="BAA95539.1"
                /db_xref="GI:7768758"
                /translation="MGEAGAAEEACRHMGTKEEFVKVRKKDLERLTTEVMQIRDFLPR
                ILNGEVLESFQKLKIVEKNLERKEQELEQLKMDCEHFKARLETVQADNIREKKEKLAL
                RQQLNEAKQQLLQQAEYCTEMGAAACTLLWGVSSSEEVVKAILGGDKALKFFSITGQT
                MESFVKSLDGDVQELDSDESQFVFALAGIVTNVAAIACGREFLVNSSRVLLDTILQLL
                GDLKPGQCTKLKVYG"
exon            complement(7441. .7566)
                /gene="HSF2BP"
                /number=7
repeat_region   7702. .7999
                /note="AluJo"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
repeat_region   8247. .8455
                /note="MER46A"
                /rpt_family="DNA/MER2_type"
                /rpt_type=DISPERSED
repeat_region   9535. .9640
                /note="MLT1G"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
repeat_region   9714. .9959
                /note="MLT1F"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
repeat_region   9981. .10277
                /note="AluSx"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
repeat_region   10278. .10575
                /note="AluY"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
repeat_region   10708. .10829
                /note="MLT1H"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
repeat_region   11291. .11591
                /note="AluSx"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
repeat_region   11596. .11892
                /note="AluJo"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
repeat_region   12004. .12273
                /note="L1M4"
                /rpt_family="LINE/L1"
                /rpt_type=DISPERSED
repeat_region   12274. .12590
                /note="AluY"

alignment_scores:
           Quality:  102.00            Length:    326
             Ratio:  0.823               Gaps:     20
 Percent Similarity:  38.037   Percent Identity: 23.926

alignment_block:
 US-09-674-779-2 x AP001752/rev   ..

 Align seg 1/1  to reverse of: AP001752  from: 1  to: 340000

     16 MetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSerPr 32
        |||||||:::|||||||::::::         |||      ::::::|||
 211329 ATGCTGCTAGCGTGCACTGGCACGCATTCTACGTGTACCGAAACCAGCAG 211280

     32 oIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisH 49
          :::                                            |||
 211279 GCTA........................................CACT 211272

     49 isThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHis 65
              :::|||   ::::::::::|||:::   |||          :::
 211271 CCTGCTCTGATCACAAACACACTGCCTCGCGGAGGGGG.........CGG 211231

     66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValAr 82
        :::   |||||||||:::::::::
 211230 ACACAATGCTGGCTGCAGCTGAGACGCCCACCG..... 211198

     82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetS 99
         |||||||||           |||||||||::::::|||       :
 211197 .TACCAGGCCTTG......CCGAGACTGGGATCCTGGCTGTCCCACCCCA 211155

     99 erGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu... 114
        ::       :::::::::   |||         ||||||::::::
 211154 AGCCCACAGGGTCGGTGGGCCTGTCTGTGACACTATGTGGGAGGAGGGCG 211105

    115 ProTyrGlnLeuPro.....................ProGluHisLe 123
        |||      ||||||                        |||
 211104 CCAGGCTGGCTACCCAGTCTAGGGTGGGACCACAGCACTGAGGAGAAGCC 211055

    123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
         ||||||        |||         |||  |||  |||:::|||:
 211054 CTGGGGA......TGCCCACTGAGCCTCCTGGTACAGCCCCTCAGAAGCC 211011

    140 rgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyr..... 153
        ::         ||||||||||||||||      :::   :::
 211010 AC......TGCCCAGCTAACACACAGCCACCCGGTTTCTGGGGCGGAGCC 210967

    154 .........................ArgAsnProGluLeuAsnGln.. 160
                                 |||:::||||||      |||
 210966 AGGCCTGGGGCCAGTGCCCTCCTGAGCCGGCACCCAGAGGGCATGCAAAG 210917

    160 ..................................................... 160

 210916 AGGCCCCAGGGCTGGGCCCCAGGAGAAACACTGACCCTCACAGCCCAGAG 210867

    161 ........CysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSer 174
              |||:::|||||||||||||:::|||         :::::::::
 210866 GAGGCCTTTGTGTGGGAGGAGCTGCCCTGAGCTGCCTTCGAGCCTCGAAA 210817

    175 AlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeu.. 190
        ::::::   :::|||              ::::::   ||||||
 210816 GGTGTCACCCTCTGG............GGCGTCAGACCACAGGCAGCCCC 210779

    191 .....................TyrGluLeuGlnAsnArgLeuCysGln. 199
                              :::   ||||||:::      |||
 210778 AAGTGTGCTAAATGTGCTTCCCTGGCCCCTCCAGCATCCTCCCTGTGGTC 210729
```

```
   200 .....................................TyrTrpLeuGlu 203
                                              |||
210728 TGTCCTGCCACGCCTGCCTGCTGGTTCCGGCCGGGTGACAGTGGGCACCT 210679

   204 ........................HisGlyGluAsn..... 207
                               ||||||||||:::
210678 GCTGAGCGGGTCTCTGCAGCCACCCTGCACGGCGAACGGCCCGGCCAGTT 210629

   208 ..........................GlnAsnPheGlyLeuGlyLeuT 215
                                |||||   |||||||||   :
210628 GAGGTCTCCAGTGGCCTCGCTGTGACCTCAGAACCAGGGACTGGGT...C 210582

   215 yrAlaThrGly.AlaIleHisLeuAspThrGlnGlyPheArgLysTrpGl 231
       ::   :::||| |||          |||||||||              ||
210581 ACAGGTCAGGTTGCACCTCCTCGACCCACACAAGGC...........GG 210544

   231 y.....................AlaGlnPheSerGluThrAsnSerIleC 241
       |                    ||||||   |||      :::|||   |
210543 CCACCTGCCCCAAGCCAAGTGTGCCCAGACCTCCGTCCTGAGCTCAGCCT 210494

   241 ys.........ArgHisValLeuPro 246
       ||         |||   :::||||||
210493 GCAGCAGATGGCGCTCTCTGCTGCCC 210468
```

seq_name: gb_pr:AC078925

seq_documentation_block:

LOCUS AC078925 153284 bp DNA linear PRI 03-NOV-2001
DEFINITION Homo sapiens 12q BAC RP11-76C10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC078925
VERSION AC078925.18 GI:13470109
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 153284)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 153284)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 153284)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 153284)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 153284)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 153284)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Mar 28, 2001 this sequence version replaced gi:13374603.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found

```
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

            QUALSTAT-REPORT.
FEATURES              Location/Qualifiers
     source           1. .153284
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="12q"
                      /clone="RP11-76C10"
     repeat_region    complement(567. .695)
                      /rpt_family="MIR"
     repeat_region    1360. .1468
                      /rpt_family="MIR"
     repeat_region    1847. .2518
                      /rpt_family="MER76"
     repeat_region    2538. .2668
                      /rpt_family="MIR"
     repeat_region    complement(2964. .3173)
                      /rpt_family="L1M4"
     repeat_region    4278. .4312
                      /rpt_family="AT_rich"
     repeat_region    4734. .4918
                      /rpt_family="(CACCC)n"
     repeat_region    complement(6347. .6727)
                      /rpt_family="MSTD"
     repeat_region    6761. .6790
                      /rpt_family="(TCTA)n"
     repeat_region    6791. .7135
                      /rpt_family="MER47A"
     repeat_region    7136. .7286
                      /rpt_family="(TCTA)n"
     repeat_region    7192. .7334
                      /rpt_family="(TCTCTG)n"
     repeat_region    7558. .7942
                      /rpt_family="MLT1B"
     repeat_region    complement(8748. .8955)
                      /rpt_family="MER58A"
     repeat_region    complement(8973. .9274)
                      /rpt_family="AluY"
     repeat_region    9407. .9469
                      /rpt_family="MER5B"
     repeat_region    9470. .9770
                      /rpt_family="AluSx"
     repeat_region    9771. .9809
                      /rpt_family="MER5B"
     repeat_region    10871. .10989
                      /rpt_family="MLT1K"
     repeat_region    11135. .11440
                      /rpt_family="AluSx"
     repeat_region    complement(12720. .12874)
                      /rpt_family="MIR"
     repeat_region    14470. .14522
                      /rpt_family="CT-rich"
     repeat_region    14766. .14958
                      /rpt_family="MER20"
     repeat_region    15105. .15523
                      /rpt_family="MLT1K"
     repeat_region    15743. .15830
                      /rpt_family="MER103"
     repeat_region    16797. .16960
                      /rpt_family="MIR"
     repeat_region    17612. .17718
                      /rpt_family="MLT1I"
     repeat_region    18796. .18824
                      /rpt_family="(TG)n"
     repeat_region    19477. .21422
                      /rpt_family="L1M4"
     repeat_region    21423. .21721
                      /rpt_family="AluJb"
     repeat_region    21722. .22069
                      /rpt_family="L1M4"
     repeat_region    22113. .23640
                      /rpt_family="L1MEc"
     repeat_region    23781. .23951
                      /rpt_family="(CATCC)n"
     repeat_region    24250. .24458
                      /rpt_family="MLT1K"
     repeat_region    25808. .25901
                      /rpt_family="MER5A"
     repeat_region    26552. .26645

alignment_scores:
            Quality:  101.50              Length:    286
              Ratio:  0.787                 Gaps:     16
 Percent Similarity:  45.105   Percent Identity: 25.524

alignment_block:
 US-09-674-779-2 x AC078925   ..

 Align seg 1/1  to: AC078925  from: 1  to: 153284

      21 SerAlaProIleProThr...AsnProGlnValSerProIleLysThrPr 36
         |||:::|||||||||:::   :::|||      ||||||      |||||
  117145 TCATCTCCAATCCCCTCTCCATCTCCATTCCCATCTCCATCTCCAACTCC 117194

      36 oSerVal.LeuIleThrLysAspLysIleGlyAspHisHisThrHisGlu 52
         ||||    |||      ::::::  :::          |||   |||
  117195 ATCTCCATCTCCAACTCCAGTCCCCTCTC.........CATCTCCATCTC 117235

      53 HisAspGluSerValSerHisValGlyLeuGlnAla......HisPheGl 67
         |||::::::       |||:::  ||||||:::      |||:::::
  117236 CATTCCCATCTCCATCTCCATCTCCAACTCCAGTCCCCTCTCCATCTCCA 117285

      67 uThrTrpLeuGlnMetHisHisAlaThrLys..................... 77
         :      |||   :::||||||      :::
  117286 TTTCCATCTCTATCTCCATCATCTCCAGTCCCCTCTCCATCTCCATTTCC 117335

      78 ...................GlnGluValValArgTyrGlnAlaTyrLeu 87
                            |||:::   ::::::   :::   |||:::
  117336 ATCTCTATCTCCAACATCTCCAGTCCCCTCTCCATCTCCATCTCTATATC 117385

      88 GlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrTh 104
         ||||||||||||      :::|||           :::|||
  117386 CAGTCCCGTCTCCATCTCCATCTCCATTCCCATGTCCATCTCAATCTCCA 117435

     104 rAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProG 121
           :::|||                                    |||
  117436 ATTCCAGTCC.....................................CCTC 117449

     121 luHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu 137
          ||||||   :::::::::   |||   |||||||||   :::   |||
  117450 TGCATCTCCATTCCCATCTCCATCCGCACCTCCATCTCCAGTCCCCTCTC 117499

     138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
           :::      ||||||   ::::::::::::|||      |||::
  117500 CATCTCCAT......CTTCCACTCCAGTCCCATCTCCGTCTCCAGTACCA 117543

     154 g.Asn......ProGluLeuAsnGln.........CysAlaGly...... 163
         : :::      |||:::|||:::|||         |||:::|||
  117544 TCTCTTGACTGCCCCAGCTCACCCAGCAGCACAAGTGTGTGGGCCCCTCC 117593

     164 ..........................GlyAlaAlaMetSerLysHis.. 170
                                    |||       |||   |||
  117594 CAGGTGGGGCAGAAACACAGTCACCCTGGAAAGAACAGAAGCACCCACCC 117643

     171 .......LeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluI 185
                |||:::||||||:::   |||         :::   |||
  117644 AGGTATGCGCACAGAGTCGGCCTTGGCCATAGAGCAGAGTGAAACAGAAT 117693

     185 leLysSerGlnAlaLeuTyrGlu.......................... 192
           :::   ::::::      |||
  117694 GAAATAGAAAATCCTGCCAGGAAATGCCTTTAGTGTTAATTCTGACCAAT 117743

     193 ..........................LeuGlnAsn.ArgLeuCysGlnT 200
                                    |||:::|||    |||   |||
```

```
  117744 AACACAGTGAGTTCTTGAATAAACACTTTGAAAAATCTCACTCTTTCAGT 117793

     200 yrTrpLeuGluHisGlyGluAsn...GlnAsnPheGlyLeuGlyLeuTyr 215
         ::::::|||     :::   ::::::   |||:::|||
  117794 GTCACATCAAACACCATACCTCCTGACATGAACCTGGCGTGGGGAACAGC 117843

     216 AlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAl 232
         |||                      :::   |||   :::   :::||
  117844 GCA.....................CGCGTTTTTTAAAGAACGGCAGC 117869

     232 aGlnPheSerGluThrAsnSerIleCysArgHisValLeuProLysAsnL 249
         |   :::|||      :::|||   |||   |||:::||||||   :::|
  117870 ATTATACTCTGGTTTCTCCTCAGATTGTATACATCTTTTGCCTTTTACTA 117919

     249 ys 249
         ||
  117920 AA 117921
```

seq_name: gb_pr:AL133414

seq_documentation_block:

LOCUS AL133414 159169 bp DNA linear PRI 01-JUN-2001

DEFINITION Human DNA sequence from clone RP5-1060P11 on chromosome 19. Contains ESTs, STSs and GSSs. Contains 10 natural killer cell receptors, including the KIR2DL2, KIR2DL4, KIR2DS1 and KIR2DS2 genes for killer cell two domain long cytoplasmic tail immunoglobulin-like receptor 2 and 4 and and short cytoplasmic tail immunoglobulin-like receptor 1 and 2 and the KIR3DL2 and KIR3DS1 genes for killer cell three domain short cytoplasmic tail immunoglobulin-like receptor 2 and short cytoplasmic tail immunoglobulin-like receptor 1, complete sequence.

ACCESSION AL133414

VERSION AL133414.4 GI:14329489

KEYWORDS HTG; KIR2DL2; KIR2DL4; KIR2DL5; KIR2DS1; KIR2DS2; KIR3DL2; KIR3DS1; NKAT.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 159169)

AUTHORS Milne,S.

TITLE Direct Submission

JOURNAL Submitted (26-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Jun 8, 2001 this sequence version replaced gi:7242326.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated as part of a collaboration with Michael Wilson, Michaela Torkar and John Trowsdale, Department of Pathology, Division of Immunology, the University of Cambridge, UK.
RP5-1060P11 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-1060P11 The true right end of clone RP1-52N12 is at 11343 in this sequence.

FEATURES Location/Qualifiers

| | |
|---|---|
| source | 1. .159169 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /chromosome="19" |
| | /clone="XX-1060P11" |
| misc_feature | complement(1. .156) |
| | /note="match: GSS: Em:AQ218339" |
| repeat_region | 100. .395 |
| | /note="L1M4 repeat: matches 3275. .3598 of consensus" |
| repeat_region | 468. .4510 |
| | /note="L1PA16 repeat: matches 2149. .6156 of consensus" |
| misc_feature | complement(4308. .4835) |
| | /note="match: GSS: Em:AQ680026" |
| repeat_region | 4834. .5143 |
| | /note="AluSx repeat: matches 3. .312 of consensus" |
| repeat_region | 5145. .5246 |
| | /note="L1M4c repeat: matches 1766. .1869 of consensus" |
| repeat_region | 5667. .5916 |
| | /note="L1M4 repeat: matches 2363. .2622 of consensus" |
| repeat_region | 5975. .6067 |
| | /note="FLAM_C repeat: matches 26. .118 of consensus" |
| repeat_region | 6107. .6555 |
| | /note="L1M4 repeat: matches 2637. .3130 of consensus" |
| repeat_region | 6701. .7062 |
| | /note="L1ME3 repeat: matches 5571. .5927 of consensus" |
| repeat_region | 7075. .7104 |
| | /note="15 copies 2 mer gt 96% conserved" |
| repeat_region | 7155. .7449 |
| | /note="AluSx repeat: matches 1. .295 of consensus" |
| repeat_region | 7494. .7566 |
| | /note="FLAM repeat: matches 46. .103 of consensus" |
| gene | join(7915. .7948,8626. .8661,9432. .9716,10990. .11289, 12868. .13161,18629. .18733,19196. .19248,19347. .19472) |
| | /gene="1060P11.1" |
| CDS | join(7915. .7948,8626. .8661,9432. .9716,10990. .11289, 12868. .13161,18629. .18733,19196. .19248,19347. .19472) |
| | /gene="1060P11.1" |
| | /note="match: proteins: Tr:O95056" |
| | /codon_start=1 |
| | /evidence=not_experimental |
| | /product="1060P11.1 (killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3 (KIR3DL3) (KIRCI, KIR411))" |
| | /protein_id="CAC40702.1" |
| | /db_xref="GI:14329490" |
| | /translation="MSLMVVSMACVGFFLLEGPWPHVGGQDKPFLSAWPGTVVSEGQH VTLQCRSRLGFNEFSLSKEDGMPVPELYNRIFRNSFLMGPVTPAHAGTYRCCSSHPHS PTGWSAPSNPVVIMVTGVHRKPSLLAHPGPLVKSGETVILQCWSDVRFERFLLHREGI TEDPLRLIGQLHDAGSQVNYSMGPMTPALAGTYRCFGSVTHLPYELSAPSDPLDIVVV GLYGKPSLSAQPGPTVQAGENVTLSCSSRSLFDIYHLSREAEAGELRLTAVLRVNGTF QANFPLGPVTHGGNYRCFGSFRALPHAWSDPSDPLPVSVTGNSRYLHALIGTSVVIIP FAILLFFLLHRWCANKKNAVVMDQEPAGNRTVNREDSDEQDPQEVTYAQLNHCVFTQR KITRPSQRPKTPPTDTSV" |
| repeat_region | 7991. .8597 |
| | /note="L1M3e repeat: matches -539. .-70 of consensus" |
| misc_feature | 9532. .10003 |
| | /gene="1060P11.1" |
| | /note="match: GSS: Em:AQ353217" |
| repeat_region | 9859. .10356 |
| | /note="MLT1D repeat: matches 1. .505 of consensus" |
| repeat_region | 12157. .12214 |
| | /note="29 copies 2 mer ga 70% conserved" |
| misc_feature | complement(12198. .12445) |
| | /note="match: GSS: Em:AI001046" |
| misc_feature | 12201. .12325 |
| | /gene="1060P11.1" |
| | /note="match: STS: Em:G10334" |
| misc_feature | 12321. .12464 |
| | /gene="1060P11.1" |
| | /note="match: STS: Em:G08241" |
| misc_feature | complement(12350. .12552) |
| | /note="match: STS: Em:G10066" |

```
repeat_region   12352. .12437
                /note="43 copies 2 mer at 66% conserved"
misc_feature    complement(12420. .12559)
                /note="match: GSS: Em:AF017584"
repeat_region   12684. .12767
                /note="42 copies 2 mer ag 67% conserved"
repeat_region   13563. .13646
                /note="L1MA4 repeat: matches 6081. .6161 of consensus"
repeat_region   13647. .13952
                /note="AluSx repeat: matches 1. .286 of consensus"
repeat_region   13953. .14375
                /note="L1MA4 repeat: matches 5606. .6081 of consensus"
repeat_region   14376. .14680
                /note="AluSq repeat: matches 1. .308 of consensus"
repeat_region   14681. .14764
                /note="L1MA4 repeat: matches 5526. .5606 of consensus"
repeat_region   14898. .15300
                /note="MLT1D repeat: matches 123. .497 of consensus"
repeat_region   15348. .15461
                /note="MLT1D repeat: matches 1. .105 of consensus"
repeat_region   15622. .15931
                /note="AluSx repeat: matches 1. .306 of consensus"
repeat_region   16636. .16914
                /note="AluSx repeat: matches 11. .299 of consensus"
misc_feature    17494. .17951
                /gene="1060P11.1"
                /note="match: GSS: Em:AQ878926"
repeat_region   17670. .17998
                /note="MER70A repeat: matches 147. .490 of consensus"
misc_feature    17958. .18052
                /gene="1060P11.1"
                /note="match: GSS: Em:AQ878926"
repeat_region   17999. .18435
                /note="MSTB repeat: matches 6. .426 of consensus"
repeat_region   19667. .19927
                /note="L2 repeat: matches 2152. .2419 of consensus"
repeat_region   20527. .20608
                /note="MER2 repeat: matches 106. .170 of consensus"
repeat_region   20609. .20912
                /note="AluSg repeat: matches 1. .291 of consensus"
repeat_region   20913. .21007
                /note="MER2 repeat: matches 6. .106 of consensus"
repeat_region   21092. .21356
                /note="L1M4 repeat: matches 4699. .4976 of consensus"
repeat_region   21370. .21488
                /note="AluSg/x repeat: matches 184. .302 of consensus"
mRNA            join(21940. .21973,22882. .22917,25354. .25653,27168. .27461,
                30731. .30781,35047. .35148,35611. .35663,35762. .36449)
                /gene="KIR2DL2"
                /note="killer cell immunoglobulin-like receptor, two
                domains, long cytoplasmic tail, 1 (NKAT1)"
                /product="1060P11.2 (killer cell immunoglobulin-like
                receptor, two domains, long cytoplasmic tail, 2 (NKAT6))"
                /evidence=not_experimental
gene            21940. .36449
                /gene="KIR2DL2"
CDS             join(21940. .21973,22882. .22917,25354. .25653,27168. .27461,
                30731. .30781,35047. .35148,35611. .35663,35762. .35938)
                /gene="KIR2DL2"
                /note="match: proteins: Sw:P43627"
                /codon_start=1
                /evidence=not_experimental
                /product="1060P11.2 (killer cell immunoglobulin-like
                receptor, two domains, long cytoplasmic tail, 2 (NKAT6))"
                /protein_id="CAC40703.1"
                /db_xref="GI:14329491"
                /translation="MSLMVVSMACVGFFLLQGAWPHEGVHRKPSLLAHPGRLVKSEET
                VILQCWSDVRFEHFLLHREGKFKDTLHLIGEHHDGVSKANFSIGPMMQDLAGTYRCYG
                SVTHSPYQLSAPSDPLDIVITGLYEKPSLSAQPGPTVLAGESVTLSCSSRSSYDMYHL
                SREGEAHECRFSAGPKVNGTFQADFPLGPATHGGTYRCFGSFRDSPYEWSNSSDPLLV
                SVTGNPSNSWPSPTEPSSKTGNPRHLHILIGTSVVIILFILLFFLLHRWCSNKKNAAV
                MDQESAGNRTANSEDSDEQDPQEVTYTQLNHCVFTQRKITRPSQRPKTPPTDIIVYTE
                LPNAESRSKVVSCP"
```

```
repeat_region   22006. .22760
                /note="L1M3e repeat: matches -539. .867 of consensus"

alignment_scores:
            Quality:  101.50            Length:    224
              Ratio:  1.005               Gaps:     11
 Percent Similarity:  45.089   Percent Identity: 25.893

alignment_block:
 US-09-674-779-2 x AL133414/rev   ..

 Align seg 1/1  to reverse of: AL133414  from: 1  to: 159169

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
         |||   |||   :::|||      |||   |||:::||||||
   98948 CCCACGCCCATATCTCCACTCCAGTCCCATATCTCCACCCCATGCCCATA 98899

      39 uIleThrLysAspLysIleGlyAspHisHisThrHis..... 51
          :::   ::::::   |||:::      |||||||||
   98898 TCTGCACTCCAGTCCCATATCTCCACCCCACACCCATATCTCCACTTCAG 98849

      52 ........GluHisAspGluSerValSer.....HisValGly 61
                 |||::::::   :::|||           |||::::::
   98848 TCCCATATCTCCACTCAAGGCCCATATCTCCACCCCACGCCCATATCTCC 98799

      62 .LeuGlnAlaHisPheGluThr.............TrpLeuGlnMetH 73
          ||||||||||||:::::|||                ||||||   |
   98798 ACTCCAGGCCCATATCTCCACTCCAGGCCCATATCTCCACCTCCAGGCCC 98749

      73 isHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSer 89
         ||   :::|||           ::::::   ||||||::::::   |||
   98748 ATATCTCCACTCCAGGCCCATATCTCCATCTCCAGGCTCATATC...TCC 98702

      90 ArgLeuGlyAsnTyrLeu.................ProProMetSerGl 100
            |||   ||||||                ||||||   :::
   98701 ACTCCAGGCCCATATCTCCACTCCAGGCCCTTATCTCCACCTCCAGGCCC 98652

     100 nLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG 117
          |||   :::      :::          ||||||
   98651 ATATCTCCACTCCAGACCCACATCTCCACTCCAGGGCCATATCTCCACTC 98602

     117 lnLeuProProGluHisLeuTrpGlyGlnIleVal.........ProThr 130
            |||      ||||||   ::::::|||              |||
   98601 CAGGTCCATATCTGCACCTCCAGGCCCATATCTCCACTCCAGGCCCATAT 98552

     131 LeuHisLeuTyrGlnAsp.....................LeuLysSerAr 140
         |||||||||                           |||   |||||
   98551 CTCCACCTCCAGGCCCATAACTTCACTCCAGGCCCATAACTCCACTCCAG 98502

     140 g......GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgA 155
         |      ::::::|||   ::::::   :::             |||   :
   98501 GCCCATATCTCTACTCCAGTCCCATATCTCCACTCCAGTCCCATATCTCC 98452

     155 snProGluLeu.....................AsnGlnCysAlaGlyGly 164
         ::|||:::|||                   :::      |||:::
   98451 ACCCTAGGCTCCTACCTCCCATCCAGGTTCCTATCTCTTCTCCAGGTTCC 98402

     165 AlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPr 181
         :::      |||:::   ::::::               |||   :::||
   98401 TCTCTCCACTCCAGGCCCATATCTCCACTGCAGGCCCATATCTCCACTCC 98352

     181 oAspLeuGluIleLys.....................SerGlnAlaLeuT 191
         ||||   ::::::                      |||:::   :::|
   98351 AGACCCAGATCTCCACTTCTAGGCCCATCACTCCATCTCTAGGCCCATAT 98302

     191 yrGluLeuGlnAsnArgLeu 197
         ||   ||||||   ::::::
   98301 ATCCCCTCCAGGCCCAGATC 98282

seq_name: gb_pr:AC069304

seq_documentation_block:
LOCUS       AC069304               161247 bp    DNA     linear   PRI 07-NOV-2001
```

```
DEFINITION  Homo sapiens BAC clone RP11-632K21 from 7, complete sequence.
ACCESSION   AC069304
VERSION     AC069304.7  GI:13549358
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161247)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 161247)
  AUTHORS   Isak,A., Maupin,R. and Reed,J.
  TITLE     The sequence of Homo sapiens BAC clone RP11-632K21
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 161247)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 161247)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 161247)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 161247)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 5, 2001 this sequence version replaced gi:13431246.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0632K21
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP5-1051J4, 200 bp overlap.
            Actual start of this clone is at base position 152813 of
            RP5-1051J4; actual end is at base position 161247 of RP11-632K21.
FEATURES             Location/Qualifiers
     source          1. .161247
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-632K21"
                     /clone_lib="RPCI-11"
     repeat_region   40. .71
                     /rpt_family="(TAAA)n"
     repeat_region   2090. .2147
                     /rpt_family="(GAAA)n"
     repeat_region   2280. .2340
                     /rpt_family="MaLR"
     misc_feature    2381. .2572
                     /note="similar to  EST BG033953 (NID:g12426761)"
     misc_feature    2569. .2785
                     /note="similar to  EST W67711 (NID:g1376582) zd42e02.r1"
     misc_feature    2595. .2633
                     /note="similar to  EST AW007542 (NID:g5856405) wt02b11.x1"
     misc_feature    2595. .2633
                     /note="similar to  EST AL518924 (NID:g12782417)"
     misc_feature    2649. .2785
                     /note="similar to  EST AL518924 (NID:g12782417)"
     misc_feature    2652. .2785
                     /note="similar to  EST AW007542 (NID:g5856405) wt02b11.x1"
     repeat_region   2792. .2896
                     /rpt_family="L1"
     misc_feature    2887. .3237
                     /note="similar to  EST W67711 (NID:g1376582) zd42e02.r1"
     misc_feature    2887. .3280
                     /note="similar to  EST AL518924 (NID:g12782417)"
     misc_feature    2887. .3280
                     /note="similar to  EST AW007542 (NID:g5856405) wt02b11.x1"
     misc_feature    2899. .3275
                     /note="similar to  EST BG403336 (NID:g13296784)"
     misc_feature    2904. .3275
                     /note="similar to  EST BF541175 (NID:g11628556)"
     misc_feature    2961. .3275
                     /note="similar to  EST BF982899 (NID:g12385711)"
     repeat_region   4160. .4427
                     /rpt_family="L1"
     repeat_region   4429. .4832
                     /rpt_family="L1"
     repeat_region   4846. .4966
                     /rpt_family="L2"
     misc_feature    5048. .5290
                     /note="similar to  EST AI002182 (NID:g3202516) oq85c05.s1"
     repeat_region   5285. .6565
                     /rpt_family="L1"
     repeat_region   7323. .8249
                     /rpt_family="L1"
     repeat_region   9458. .9606
                     /rpt_family="MaLR"
     repeat_region   9618. .9971
                     /rpt_family="L1"
     repeat_region   9975. .10442
                     /rpt_family="L1"
     repeat_region   10523. .10828
                     /rpt_family="ERV1"
```

```
     repeat_region   10974. .11095
                     /rpt_family="L2"
     misc_feature    11933. .12225
                     /note="similar to  EST AL445235 (NID:g13396667)"
     repeat_region   11998. .12306
                     /rpt_family="Alu"
     misc_feature    12246. .12531
                     /note="similar to  EST AL445235 (NID:g13396667)"
     repeat_region   12586. .12696
                     /rpt_family="MIR"
     repeat_region   12762. .12940
                     /rpt_family="MER1_type"
     repeat_region   13270. .13451
                     /rpt_family="MER1_type"
     repeat_region   13975. .14140
                     /rpt_family="MIR"
     repeat_region   14357. .14477
                     /rpt_family="CT-rich"
     repeat_region   16741. .17304
                     /rpt_family="L1"
     misc_feature    17292. .17774
                     /note="similar to  EST AW207586 (NID:g6507082)"
     misc_feature    17292. .17733
                     /note="similar to  EST BF691430 (NID:g11976838)"
     misc_feature    17302. .17516
                     /note="similar to  EST BE150979 (NID:g8613700)"
     misc_feature    17302. .17774
                     /note="similar to  EST AI419921 (NID:g4265852) tg40c01.x1"
     misc_feature    17337. .17774
                     /note="similar to  EST AV682285 (NID:g10284148)"
     misc_feature    17341. .17758
                     /note="similar to  EST AV744496 (NID:g10862077)"
     misc_feature    17390. .17765
                     /note="similar to  EST AW594696 (NID:g7281921) hg69h12.x1"
     misc_feature    17405. .17774
                     /note="similar to  EST BF225559 (NID:g11133130)
                     uy42g01.x1"
     misc_feature    17405. .17770
                     /note="similar to  EST BF660480 (NID:g11925614)"
     misc_feature    17412. .17768
                     /note="similar to  EST BE535697 (NID:g9764342)"
     misc_feature    17421. .17762
                     /note="similar to  EST AW313747 (NID:g6743003)"
     misc_feature    17489. .17774
                     /note="similar to  EST AI124003 (NID:g3539769) qa99e07.s1"
     misc_feature    17544. .17768
                     /note="similar to  EST AI253794 (NID:g3858717) aq30c05.x1"
     misc_feature    17547. .17774
                     /note="similar to  EST AA897328 (NID:g3033948) al47e06.s1"
     misc_feature    17583. .17770
                     /note="similar to  EST D19777 (NID:g500093)"
     repeat_region   17913. .18030
                     /rpt_family="L2"

alignment_scores:
           Quality:  101.50            Length:    207
             Ratio:  0.949               Gaps:     13
 Percent Similarity:  51.691   Percent Identity: 25.121

alignment_block:
 US-09-674-779-2 x AC069304/rev   ..

 Align seg 1/1  to reverse of: AC069304  from: 1  to: 161247

      42 LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSe 58
         ||||||:::        :::     |||:::||||||   :::::::||
  141038 AAGGACAGAACCAAGCCACAAGGGATTCACTCCCATGACCCAAACACTTC 140989

      58 rHis.................ValGlyLeuGlnAlaHisPheGluThrT 69
         ||||                 :::|||:::      :::   |||   |
  140988 TCACCAGACCCCACTTCCAACATTGGGGATTATAATTCAACATGAGATTT 140939

      69 rp............LeuGlnMetHisHisAlaThrLysGlnGluValVal 81
         ||            :::::::::::::|||:::|||         :::
  140938 GGGAGGGACATATATTTAAACTATATCATTCCACC...CCTGGCCTTCCA 140892

      82 ArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuPro..... 96
         ::::::   ::::::   :::   ::::::   :::   ||||||
  140891 AATTTCATATCCTTCTCACATTGTAAAATAAAATCATGCCTTCCCAACAG 140842

      97 .......ProMetSerGlnLeuLeuThrThr.AlaArgSerTrpGlnAla 110
                ||||   :::|||:::|||||||::: :::   |||   |||
  140841 CCCCCTGCCCCACACACAAATCTTAACTAATTTCAGCATCAACTCAA... 140795

     111 CysGlyHisGluPro.TyrGlnLeuProProGluHisLeuTrpGlyGlnI 127
            |||       |||    |||. |||   :::||||||         :
  140794 ...GGAAAAGTTCCAAGTCCAAAGTCCAAAGTCTCATCTG.........G 140757

     127 leValProThrLeuHisLeu...TyrGlnAspLeuLysSerArgGlyIle 142
         :::::   |||:::||||||             :::|||:::      :::
  140756 TAATGAGTACCTTCCACCTATGAGCTTGTAAAATCAAAACAAGTTATTTA 140707

     143 LeuProAlaAsnThrGlnIleArgSerVal.........TyrArgAsnPr 156
         ||||||      :::   :::   ::::::           :::|||
  140706 CTTCCAAGATACAATGGGGTATAGGCATTGGGTAAACCACTTCAAAACAA 140657

     156 oGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrA 173
          |||:::::::::                          |||      |||:
  140656 AGAAATCAGCCAC.......................AAGAAAGAGACTA 140631

     173 snSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAla 189
         :::::   |||::::::|||   |||||||||       ||||||
  140630 CAGGCCTCATACAAGTTTGGAACCCAGACTTGTATGGGAAGTCATTAAAT 140581

     190 LeuTyrGluLeuGlnAsnArg............LeuCysGlnTyrTrpLe 202
         |||:::   :::   :::                |||      |||
  140580 CTGTTTGTTTTTTTCAAGATGGGGATCTCATTTTGTTGCCCAGGTTGGTC 140531

     202 uGlu...HisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrG 218
          :::      |||   :::|||:::      |||||||||      :::|
  140530 TCAAACTCCTGGCCTCAAGCAATCCTCCTACCTCGGCCTCCCAAATGCTG 140481

     218 lyAlaIleHisLeuAsp 223
         ||      :::::::::
  140480 GGATTACACGCATGAAC 140464

seq_name: gb_ba:AE008307

seq_documentation_block:
LOCUS       AE008307              11814 bp    DNA     linear   BCT 18-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 111
            of 187 of the complete sequence.
ACCESSION   AE008307 AE007870
VERSION     AE008307.1  GI:15159536
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (Cereon).
  ORGANISM  Agrobacterium tumefaciens str. C58 (Cereon)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 11814)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11814)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Approximately 800 bp of telomeric sequence missing from the left
            end of the chromosome and 200 bp missing from the right end.
FEATURES             Location/Qualifiers
     source          1. .11814
                     /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
                     /strain="C58"
```

```
                     /db_xref="taxon:181661"
gene                 complement(145. .519)
                     /gene="AGR_L_2135"
CDS                  complement(145. .519)
                     /gene="AGR_L_2135"
                     /note="(L27995) outer membrane lipoprotein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2135p"
                     /protein_id="AAK89638.1"
                     /db_xref="GI:15159537"
                     /translation="MNFKTSAALLSAAIAVSSCVAPGPSQQTSMAPSLSRGPAVDGRW
                     IDRNGIISTFQNGAFSTRSTDTNTLLASGTYVTISPTLYEINMTSLVRNTQSRVNCAL
                     VSPSQLNCTTDANSQFSLTRQG"
gene                 complement(669. .2114)
                     /gene="AGR_L_2136"
CDS                  complement(669. .2114)
                     /gene="AGR_L_2136"
                     /note="(L77975) homospermidine synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2136p"
                     /protein_id="AAK89639.1"
                     /db_xref="GI:15159538"
                     /translation="MTDANYPIYGDIDGPIIMIGFGSIGRGTLPLIERHFNFDRTRLV
                     VIDPREDIAAFLAERNIRHVQTHLTKENYKDVLKPLLKGVQGKGFCVNLSVDAASVDL
                     MRICRKHGMLYIDTVVEPWPGFYFDADADNRARTNYSLRETMLKERARKPGGTTAVST
                     CGANPGMVSWFVKQALVNLAKDTGLDIEEPQSHDREGWAKLMQQLGVKGVHVAERDTQ
                     RAKEPKPFGAFWNTWSVEGFIAEGFQPAELGWGTHENWMPKNAKKQKKGSKAAIYLDQ
                     PGANTRVRTWCPGLGAQYGFLVTHNEAISISDYFTVRNDDGDVTYRPTCHYAYHPCND
                     AILSLHELFGNGGNPQPIQHVLGEDELVDGADELGVLLYGHDRNAYWYGSRLSLEEAR
                     SLAPSQNATGLQVSSAVLAGMAWAIENPEAGIVEADEMDYRRCLEVQRPYLGPVEGHY
                     TDWTPLDGRPGLFPDNIDMSDPWQFRNILVR"
gene                 2374. .3282
                     /gene="AGR_L_2138"
CDS                  2374. .3282
                     /gene="AGR_L_2138"
                     /note="hypothetical protein PA0290 [imported] -
                     Pseudomonas aeruginosa (strain PAO1)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2138p"
                     /protein_id="AAK89640.1"
                     /db_xref="GI:15159539"
                     /translation="MNDLLERVAALQSATPVLIALYDSNDRLRHANPSFRESFQLADG
                     DFPTWADIMRLNHRLRTGVVIRNNDFEGWLLSAQSRRGKLPYRAFEMDLYDGRWFWMT
                     ETVQQDGWMLCVASDVTELRQGYRTLRQDRDLAVRASQTDELTGTANRRFVFAKLALQ
                     VEQVSKGLEAPFCLCTLDIDFFKQINDRFGHKCGDAVLRDFVGIAHSVIRRVDCFGRI
                     GGEEFMLILPRSSIAEGREIVERLLQKVRLSRPLSEFPDFSYSCSAGMTVFRMMENAD
                     DVFMRADQALYAAKEKGRDRLLAVTV"
gene                 complement(3279. .3974)
                     /gene="AGR_L_2139"
CDS                  complement(3279. .3974)
                     /gene="AGR_L_2139"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2139p"
                     /protein_id="AAK89641.1"
                     /db_xref="GI:15159540"
                     /translation="MPVDIASGETGKPSRRRAQAMKRKPADLTLRETLRWEPQTGFQR
                     IDQHMRRLLRSADALGFRAPVNAVQVLEKQVGGEKPLCVRLVMNYKGELDIETVEFRP
                     LPENAVWRVRIARQTTLDSADTFYRHKSSRREPYDAARAEFSAEEADEVLLLNERGEL
                     CEGSSTSIFVEMPEGPMLTPPLDSGILPGVLRADLIRERRARGQALKPEDIAGRKIFV
                     GNSLEGLIAAELI"
gene                 complement(4014. .6089)
                     /gene="AGR_L_2140"
CDS                  complement(4014. .6089)
                     /gene="AGR_L_2140"
                     /note="oligoendopeptidase F - Helicobacter pylori (strain
                     26695)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2140p"
                     /protein_id="AAK89642.1"
                     /db_xref="GI:15159541"
                     /translation="MSRRRWTGSSSITRTCAVIFSSGFRFMDLLHAVPYAALWHGRFE
                     HPVLRIACILSIAAGNRVPICFESITRRIRPMTLHRPAPHPVFSLAETSGPVLGDLPG
                     WKLSDLYPSGDSPEYKRDLEKAASMAAAFEAKWKDKLESAAKAKGEDGIGAALKDYEA
                     LDDLIGRIGSFAGLTYFSDTSNPANGKLYGDAQAKLTEISAHLLFFTLELNRIDDAVM
                     DACMANDAAAGHYRPWLVDLRKDKPHQLDDQLEQLFLEKSMTSAAAFNRLFDETMADL
                     RFDIDGASLSLEVTLNMLQEPEPETRRKAAEALSATFKDNLRVFTLITNTLAKDKDIS
                     DRWRNFEDIADSRHLANRVERDVVDALAAAVKKAYPRLSHRYYAMKAKWLGMEQMNYW
                     DRNAPLPDTSNTIIPWDEAKDTVLSAYGNFAPEMADIARRFFDEEWIDAPARPGKAPG
                     AFAHPTVPSAHPYVLVNYLGKPRDVMTLAHELGHGVHQVLAGQQGALMCATPLTLAET
                     ASVFGEMLTFRKLLDDTKDIRERKAMLARKVEDMINTVVRQIAFYEFERKLHTARKEG
                     ELTSEQIGELWLSVQAESLGPAIRISEGYETWWTYIPHFIHSPFYVYAYAFGDCLVNS
                     LYAVYQKAEAGFQEKYFELLKAGGTKHHSELLKPFGLDATDPSFWDKGLSMIEGLIDE
                     LEELDRKQA"
gene                 6034. .7575
                     /gene="AGR_L_2141"
CDS                  6034. .7575
                     /gene="AGR_L_2141"
                     /note="(AF010585) putative sigma 54 activator"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2141p"
                     /protein_id="AAK89643.1"
                     /db_xref="GI:15159542"
                     /translation="MTAHVLVIDDDPVQRRLLKNAVERYGHLALLAENGRAGLELLKQ
                     YSGEINVIVLDLMMPEMDGLAFLKAVGELGTDVPVIVQTGQGGIDTVVQAMRAGAFDF
                     VVKPVSPERIGAAISNALKLDRKEAKARTGRRGRSNAVNFTDIVSASPAMLRVIELAQ
                     RAAQSSIPVVLEGESGVGKEMVARAIQSGGDRANKPFITVNCGAIPHNLVESILFGHE
                     KGAFTGATEKHVGKFMEADGGTLFLDEIGDLPLDVQVKLLRAVQQGEIETVGSARVQK
                     VNVRLISATNKNLIEEVKEGRFREDLYYRLNVFPITIPALRRRKEDIPHLARVFAERF
                     SAEQKLPNPVGLDASALALLTAYDWPGNIRQLENAVFRAVVLSQGSELSDSDFPQIAL
                     QLPEFSNGDDGDDIIRSPSGSSLKLAAYSPPPAAESHAVIEEPSDISTTIYRGGNLIS
                     STDDSGNIRKLAEIEEELIRFALRFYRGQMSQVARKLGIGRSTLYRKLKDYGIDPDDP
                     LRDVA"
gene                 7752. .9656
                     /gene="AGR_L_2143"
CDS                  7752. .9656
                     /gene="AGR_L_2143"
                     /note="conserved hypothetical protein VC1269 [imported] -
                     Vibrio cholerae (group O1 strain N16961)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2143p"
                     /protein_id="AAK89644.1"
                     /db_xref="GI:15159543"
                     /translation="MTEIVFQAGIELPYLHRNNALSTKIARGLIKDICTKLSARAVTF
                     ACLMLAAMPFVGVSATEAAAETRSLKLYYIHTREKAVITFKRNGKYDQKGLQELNRFL
                     RDWRRNQPTRMDPRLFDLVWEVYRRSGATDYINVVSAFRSPETNGLLRTRTKGVAEKS
                     QHMLGKAMDFYIPGVKLATLREIGMQMQIGGVGFYPTSGSPFVHMDVGGVRAWPRMSR
                     QELVRIFPKGNTLHVPSDGKPLPGYEQALADYKKRVSSSSIQVASSAGSGPASSGGGG
                     KRKTLLQALFGGGDEDEDPDSIAAPEPAERPAVARPAPQQAAPEPEPTIAVASAQALP
                     GVNDAPLPTARPAFGNQPPANAGLATALYSPARNAAQDALQAATMPTPTSAPAERQQF
                     ADLAEVSVPVPTLLGPRGMKGDAEGSILTASADGATPAATGELASIPVPANRPAVAEA
                     LLAAANTDQEGEDDLAEAQQDTLSPTVVAALEQSGQAARSQVQSAAMPTAAPFPAAVN
                     HKVTVAPAETAKAAPQPGFGDAFDLKPAINGGLSAGLPTKGSRPSKQDAVVSEQAMVG
                     GGRLTQDLISDWALNQNKGTTGRSVKAPRVVANRMLSKEMSASATSASFKPGAVAIES
                     SRFSTPVKMH"
gene                 complement(9737. .11176)
                     /gene="AGR_L_2146"
CDS                  complement(9737. .11176)
                     /gene="AGR_L_2146"
                     /note="PYRUVATE KINASE (PK)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2146p"
                     /protein_id="AAK89645.1"
                     /db_xref="GI:15159544"
                     /translation="MKRNRKVKILATLGPASSEEAMIEKLHLAGADLFRINMSHASHD
                     VMRTLIQRIRSVESRNGRPIGILADLQGPKLRVGKFAETKVDLVPGQTFTLDNNDTPG
                     DNTRVFLPHPEILEAVKPGHRLLIDDGKLHLRAEKSDGKSIVTTVVSGTRISDRKGIS
                     LPDTLLGVGVLTDKDRVDLDAVLETNEVDWVALSFVQRPEDLAEVRKISRGRVGLMSK
```

```
                     IEKPQAVERIEEIIELSDALMVARGDLGVEMPLEAVPGIQKQLTRACRRAGKPVVVAT
                     QMLESMITAPVPTRAEVSDVATAVFEGADAIMLSAESASGDYPIEAVSTMASIASTVE
                     QDPYYSNIIYAQRPQPEATGADAISLAARQIAETLKLAAIVTYTSSGTTGLRAARERP
                     QVPIIALSPIIQTARRLSVVWGLHCVVTGDASDLDDMVNRACRIVVSEGFGKPGDRII
                     ISAGVPLGTPGATNMVRIAYIGSDGQSGV"
     gene            complement(11176. .11715)
                     /gene="AGR_L_2147"
     CDS             complement(11176. .11715)
                     /gene="AGR_L_2147"
                     /note="(U87316) orf4; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_2147p"
                     /protein_id="AAK89646.1"
                     /db_xref="GI:15159545"
                     /translation="MTFFGGHAKKCVHYNHGIRMEAVTQSSLALPRHFRHIARLFSAA
                     ALFLTPFIIVESAHADFRVCNSTQNLVGVAIGYRAQDGWVSEGWWQVPSSTCATLIEG
                     ELQSRYYYLYAEDAAHGGRWTGDVNMCVAENEFKINGVDDCFARSFQRMGFKEYDTGR
                     QGSWMVQLSDTPGTQGNQN"
BASE COUNT     2368 a   3401 c   3557 g   2488 t
ORIGIN


alignment_scores:
            Quality:  101.00            Length:    169
              Ratio:  1.174               Gaps:      9
 Percent Similarity:  50.888   Percent Identity: 22.485

alignment_block:
 US-09-674-779-2 x AE008307   ..

 Align seg 1/1  to: AE008307  from: 1  to: 11814

      74 HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerAr 90
         |||:::   ::::::   |||:::   ::::::               ||
    7974 CATACACGCGAGAAGGCGGTCATCACCTTCAAG.............CG 8008

      90 gLeuGlyAsnTyr............LeuProProMetSerGlnLeuLeuT 103
         |   |||:::|||            |||      ::::::::::::|||
    8009 CAATGGCAAATACGACCAGAAGGGCCTGCAGGAACTGAACCGTTTCCTG. 8057

     103 hrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPro 119
                |||:::|||:::          :::::::|||   ::::::
    8058 ........CGTGACTGGCGCCGC......AACCAGCCGACGCGGATGGAC 8093

     120 ProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAs 136
         |||:::                          :::   |||      ::
    8094 CCGCGA.....................CTTTTCGATCTTGTCTGGGA 8119

     136 pLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValT 153
         ::::      |||:::                    ::::::   |||::::
    8120 AGTCTATCGCCGCAGCGGCGCAACCGATTACATCAATGTCGTTTCGGCCT 8169

     153 yrArgAsnProGluLeuAsn..............GlnCysAlaGlyGly 164
         ::|||:::||||||   |||               :::      |||
    8170 TCCGATCGCCTGAGACCAATGGCCTGCTGCGCACGCGCACCAAGGGTGTC 8219

     165 AlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPr 181
         |||      |||:::|||:::   ::::::|||:::|||::::::::::||
    8220 GCGGAAAAAAGCCAGCATATGCTCGGCAAGGCCATGGATTTTTACATCCC 8269

     181 oAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuC 198
         |          :::|||      :::|||   |||:::      ::::::
    8270 CGGC......GTAAAGCTCGCCACCCTGCGTGAAATCGGCATGCAGATG. 8312

     198 ysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu 214
                                     |||      |||:::|||:::
    8313 ............................CAGATCGGCGGCGTCGGCTTT 8333

     215 TyrAlaThrGlyAla.........IleHisLeuAspThrGlnGlyPheAr 228
         |||   |||::::::         :::|||:::|||:::   |||   ||
    8334 TATCCGACCTCCGGTTCTCCCTTCGTGCATATGGATGTCGGCGGCGTTCG 8383

     228 gLysTrp 230
         |   |||
    8384 CGCCTGG 8390

seq_name: gb_ba:AE009307

seq_documentation_block:
LOCUS       AE009307              11913 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 77 of
            187 of the complete sequence.
ACCESSION   AE009307 AE008689
VERSION     AE009307.1  GI:17742189
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58 (U. Washington).
  ORGANISM  Agrobacterium tumefaciens str. C58 (U. Washington)
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 11913)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     The Genome of the Natural Genetic Engineer Agrobacterium
            tumefaciens C58
  JOURNAL   Science 294 (5550), 2317-2323 (2001)
   PUBMED   11743193
REFERENCE   2  (bases 1 to 11913)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Department of Microbiology, University of
            Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
            98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .11913
                     /organism="Agrobacterium tumefaciens str. C58 (U.
                     Washington)"
                     /strain="C58"
                     /db_xref="taxon:180835"
     gene            complement(71. .1834)
                     /gene="Atu3763"
     CDS             complement(71. .1834)
                     /gene="Atu3763"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAL44573.1"
                     /db_xref="GI:17742190"
                     /translation="MLAAMPFVGVSATEAAAETRSLKLYYIHTREKAVITFKRNGKYD
                     QKGLQELNRFLRDWRRNQPTRMDPRLFDLVWEVYRRSGATDYINVVSAFRSPETNGLL
                     RTRTKGVAEKSQHMLGKAMDFYIPGVKLATLREIGMQMQIGGVGFYPTSGSPFVHMDV
                     GGVRAWPRMSRQELVRIFPKGNTLHVPSDGKPLPGYEQALADYKKRVSSSSIQVASSA
                     GSGPASSGGGGKRKTLLQALFGGGDEDEDPDSIAAPEPAERPAVARPAPQQAAPEPEP
                     TIAVASAQALPGVNDAPLPTARPAFGNQPPANAGLATALYSPARNAAQDALQAATMPT
                     PTSAPAERQQFADLAEVSVPVPTLLGPRGMKGDAEGSILTASADGATPAATGELASIP
                     VPANRPAVAEALLAAANTDQEGEDDLAEAQQDTLSPTVVAALEQSGQAARSQVQSAAM
                     PTAAPFPAAVNHKVTVAPAETAKAAPQPGFGDAFDLKPAINGGLSAGLPTKGSRPSKQ
```

DAVVSEQAMVGGGRLTQDLISDWALNQNKGTTGRSVKAPRVVANRMLSKEMSASATSA
SFKPGAVAIESSRFSTPVKMH"
gene complement(2152. .3693)
/gene="Atu3764"
CDS complement(2152. .3693)
/gene="Atu3764"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="sigma 54 dependent transcriptional activator"
/protein_id="AAL44574.1"
/db_xref="GI:17742191"
/translation="MTAHVLVIDDDPVQRRLLKNAVERYGHLALLAENGRAGLELLKQ
YSGEINVIVLDLMMPEMDGLAFLKAVGELGTDVPVIVQTGQGGIDTVVQAMRAGAFDF
VVKPVSPERIGAAISNALKLDRKEAKARTGRRGRSNAVNFTDIVSASPAMLRVIELAQ
RAAQSSIPVVLEGESGVGKEMVARAIQSGGDRANKPFITVNCGAIPHNLVESILFGHE
KGAFTGATEKHVGKFMEADGGTLFLDEIGDLPLDVQVKLLRAVQQGEIETVGSARVQK
VNVRLISATNKNLIEEVKEGRFREDLYYRLNVFPITIPALRRRKEDIPHLARVFAERF
SAEQKLPNPVGLDASALALLTAYDWPGNIRQLENAVFRAVVLSQGSELSDSDFPQIAL
QLPEFSNGDDGDDIIRSPSGSSLKLAAYSPPPAAESHAVIEEPSDISTTIYRGGNLIS
STDDSGNIRKLAEIEEELIRFALRFYRGQMSQVARKLGIGRSTLYRKLKDYGIDPDDP
LRDVA"
gene 3863. .5713
/gene="pepF"
/note="Atu3765"
CDS 3863. .5713
/gene="pepF"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="oligoendopeptidase F"
/protein_id="AAL44575.1"
/db_xref="GI:17742192"
/translation="MTLHRPAPHPVFSLAETSGPVLGDLPGWKLSDLYPSGDSPEYKR
DLEKAASMAAAFEAKWKDKLESAAKAKGEDGIGAALKDYEALDDLIGRIGSFAGLTYF
SDTSNPANGKLYGDAQAKLTEISAHLLFFTLELNRIDDAVMDACMANDAAAGHYRPWL
VDLRKDKPHQLDDQLEQLFLEKSMTSAAAFNRLFDETMADLRFDIDGASLSLEVTLNM
LQEPEPETRRKAAEALSATFKDNLRVFTLITNTLAKDKDISDRWRNFEDIADSRHLAN
RVERDVVDALAAAVKKAYPRLSHRYYAMKAKWLGMEQMNYWDRNAPLPDTSNTIIPWD
EAKDTVLSAYGNFAPEMADIARRFFDEEWIDAPARPGKAPGAFAHPTVPSAHPYVLVN
YLGKPRDVMTLAHELGHGVHQVLAGQQGALMCATPLTLAETASVFGEMLTFRKLLDDT
KDIRERKAMLARKVEDMINTVVRQIAFYEFERKLHTARKEGELTSEQIGELWLSVQAE
SLGPAIRISEGYETWWTYIPHFIHSPFYVYAYAFGDCLVNSLYAVYQKAEAGFQEKYF
ELLKAGGTKHHSELLKPFGLDATDPSFWDKGLSMIEGLIDELEELDRKQA"
gene 5813. .6448
/gene="Atu3766"
CDS 5813. .6448
/gene="Atu3766"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL44576.1"
/db_xref="GI:17742193"
/translation="MKRKPADLTLRETLRWEPQTGFQRIDQHMRRLLRSADALGFRAP
VNAVQVLEKQVGGEKPLCVRLVMNYKGELDIETVEFRPLPENAVWRVRIARQTTLDSA
DTFYRHKSSRREPYDAARAEFSAEEADEVLLLNERGELCEGSSTSIFVEMPEGPMLTP
PLDSGILPGVLRADLIRERRARGQALKPEDIAGRKIFVGNSLEGLIAAELI"
gene complement(6445. .7353)
/gene="Atu3767"
CDS complement(6445. .7353)
/gene="Atu3767"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="GGDEF family protein"
/protein_id="AAL44577.1"
/db_xref="GI:17742194"
/translation="MNDLLERVAALQSATPVLIALYDSNDRLRHANPSFRESFQLADG
DFPTWADIMRLNHRLRTGVVIRNNDFEGWLLSAQSRRGKLPYRAFEMDLYDGRWFWMT
ETVQQDGWMLCVASDVTELRQGYRTLRQDRDLAVRASQTDELTGTANRRFVFAKLALQ
VEQVSKGLEAPFCLCTLDIDFFKQINDRFGHKCGDAVLRDFVGIAHSVIRRVDCFGRI
GGEEFMLILPRSSIAEGREIVERLLQKVRLSRPLSEFPDFSYSCSAGMTVFRMMENAD
DVFMRADQALYAAKEKGRDRLLAVTV"
gene 7613. .9058
/gene="Atu3768"
CDS 7613. .9058
/gene="Atu3768"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="homospermidine synthase"
/protein_id="AAL44578.1"
/db_xref="GI:17742195"
/translation="MTDANYPIYGDIDGPIIMIGFGSIGRGTLPLIERHFNFDRTRLV
VIDPREDIAAFLAERNIRHVQTHLTKENYKDVLKPLLKGVQGKGFCVNLSVDAASVDL
MRICRKHGMLYIDTVVEPWPGFYFDADADNRARTNYSLRETMLKERARKPGGTTAVST
CGANPGMVSWFVKQALVNLAKDTGLDIEEPQSHDREGWAKLMQQLGVKGVHVAERDTQ
RAKEPKPFGAFWNTWSVEGFIAEGFQPAELGWGTHENWMPKNAKKQKKGSKAAIYLDQ
PGANTRVRTWCPGLGAQYGFLVTHNEAISISDYFTVRNDDGDVTYRPTCHYAYHPCND
AILSLHELFGNGGNPQPIQHVLGEDELVDGADELGVLLYGHDRNAYWYGSRLSLEEAR
SLAPSQNATGLQVSSAVLAGMAWAIENPEAGIVEADEMDYRRCLEVQRPYLGPVEGHY
TDWTPLDGRPGLFPDNIDMSDPWQFRNILVR"
gene 9208. .9582
/gene="Atu3769"
CDS 9208. .9582
/gene="Atu3769"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="outer membrane lipoprotein"
/protein_id="AAL44579.1"
/db_xref="GI:17742196"
/translation="MNFKTSAALLSAAIAVSSCVAPGPSQQTSMAPSLSRGPAVDGRW
IDRNGIISTFQNGAFSTRSTDTNTLLASGTYVTISPTLYEINMTSLVRNTQSRVNCAL
VSPSQLNCTTDANSQFSLTRQG"
gene 9798. .11708
/gene="Atu3770"
CDS 9798. .11708
/gene="Atu3770"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="5'-nucleotidase"
/protein_id="AAL44580.1"
/db_xref="GI:17742197"
/translation="MKKILGLGMLSVSAITLSAGAALADYELNILHINDFHSRIESIN
KFDSTCSAEEEGKNECFGGAARLLTAINQTRDALKAGGKNVLLLNAGDNFQGSLFYTT
YKGTVEAEVLNAMKFDAMTVGNHEFDDSEDGLAGFLDKVQFPVVTANVVATAASKIGD
RVKPSIVLEVGGQKIGIVGAVANDTAELATPGPNITIAEDVAKIGEQVQKLKQDGVNK
IIALTHVGYPRDLEFIAKIPDVDVVVGGHSHTLLSNTDQKAEGPYPTLVDNPGGYKVP
VVQAGQYSKYLGDLKVVFDDNGVVKESKGDPILIDSTFKPDEATLKRIDELKAPIEAL
KSKVVGTSEGPIEGDRKVCRVKECSMGNLVADATLARVKDQGVTIAFANSGGLRSSID
GGDVSMGEVLTVLPFQNTVATFQLKGEDIRAALENGVSQIDDVAGRFMQVSGMKYSFD
RSKPAGSRIVSVDVKEGDGFVPLDPTKTYTVAANNYVRTGGDGFKVFATNAINAYDFG
PNLEDAVAAYITANSPYKPYTDGRISEVTPAGYVAPAKPAAPAPAPAAAPAPAAPAAT
APAATAPASPAPAAAAAAVSKYVVEKGDSLWKIAAEKYGDGALWHKIAKANTLKQPNH

alignment_scores:
Quality: 101.00 Length: 169
Ratio: 1.174 Gaps: 9
Percent Similarity: 50.888 Percent Identity: 22.485

alignment_block:
US-09-674-779-2 x AE009307/rev ..

Align seg 1/1 to reverse of: AE009307 from: 1 to: 11913

74 HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerAr 90

```
      |||:::   ::::::   |||:::   ::::::                ||
 1753 CATACACGCGAGAAGGCGGTCATCACCTTCAAG.............CG 1719

   90 gLeuGlyAsnTyr...........LeuProProMetSerGlnLeuLeuT 103
      |   |||:::|||           |||      ::::::::::::|||
 1718 CAATGGCAAATACGACCAGAAGGGCCTGCAGGAACTGAACCGTTTCCTG. 1670

  103 hrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPro 119
            |||:::|||:::         :::::::|||   ::::::
 1669 ........CGTGACTGGCGCCGC......AACCAGCCGACGCGGATGGAC 1634

  120 ProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAs 136
      |||:::                          :::   |||      ::
 1633 CCGCGA.....................CTTTTCGATCTTGTCTGGGA 1608

  136 pLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValT 153
      ::::      |||:::                 ::::::   |||::::
 1607 AGTCTATCGCCGCAGCGGCGCAACCGATTACATCAATGTCGTTTCGGCCT 1558

  153 yrArgAsnProGluLeuAsn.............GlnCysAlaGlyGly 164
      ::|||:::||||||   |||               :::      |||
 1557 TCCGATCGCCTGAGACCAATGGCCTGCTGCGCACGCGCACCAAGGGTGTC 1508

  165 AlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPr 181
      |||      |||:::|||:::   :::::::|||:::|||::::::::::||
 1507 GCGGAAAAAAGCCAGCATATGCTCGGCAAGGCCATGGATTTTTACATCCC 1458

  181 oAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuC 198
      |       :::|||      :::|||   |||:::      ::::::
 1457 CGGC......GTAAAGCTCGCCACCCTGCGTGAAATCGGCATGCAGATG. 1415

  198 ysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu 214
                                |||      |||:::|||:::
 1414 ............................CAGATCGGCGGCGTCGGCTTT 1394

  215 TyrAlaThrGlyAla.........IleHisLeuAspThrGlnGlyPheAr 228
      |||   |||::::::         :::|||:::|||:::   |||   ||
 1393 TATCCGACCTCCGGTTCTCCCTTCGTGCATATGGATGTCGGCGGCGTTCG 1344

  228 gLysTrp 230
      |   |||
 1343 CGCCTGG 1337
```

seq_name: gb_pr:AC006925

seq_documentation_block:

LOCUS AC006925 176186 bp DNA linear PRI 16-MAR-1999
DEFINITION Homo sapiens chromosome 17, clone hRPK.81_I_9, complete sequence.
ACCESSION AC006925
VERSION AC006925.6 GI:4417322
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 176186)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.81_I_9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176186)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 176186)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 16, 1999 this sequence version replaced gi:4406826.
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only the first 176186 base pairs of this clone are being submitted. The remainder overlaps accession number AC006600 (WICGR project L720).

FEATURES Location/Qualifiers
source 1. .176186
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hRPK.81_I_9"
/clone_lib="RPCI-11 human BAC library"
repeat_region complement(660. .747)
/rpt_family="MER103"
repeat_region 1469. .1508
/rpt_family="AT_rich"
repeat_region 3728. .3848
/rpt_family="FLAM_C"
repeat_region complement(4200. .4262)
/rpt_family="MIR"
repeat_region complement(4568. .4938)
/rpt_family="MLT1A1"
repeat_region 5512. .5556
/rpt_family="(TG)n"
repeat_region complement(5751. .6558)
/rpt_family="MER21B"
repeat_region complement(6584. .6641)
/rpt_family="MER21B"
repeat_region complement(7032. .7108)
/rpt_family="MIR"
repeat_region complement(7285. .7435)
/rpt_family="MIR"
repeat_region complement(9026. .9129)
/rpt_family="L2"
repeat_region 9741. .9774
/rpt_family="AT_rich"
repeat_region complement(10481. .10793)
/rpt_family="AluSx"
repeat_region 11659. .12188
/rpt_family="L2"
repeat_region 12546. .12744
/rpt_family="MIR"
repeat_region complement(13907. .14211)
/rpt_family="AluSc"
repeat_region 15411. .15526

```
                     /rpt_family="L2"
repeat_region        15536. .16002
                     /rpt_family="MLT1C"
repeat_region        16007. .16135
                     /rpt_family="L2"
repeat_region        16140. .16168
                     /rpt_family="(A)n"
repeat_region        complement(17311. .17680)
                     /rpt_family="THE1C"
repeat_region        complement(17681. .17823)
                     /rpt_family="THE1C-internal"
repeat_region        complement(17824. .19328)
                     /rpt_family="THE1C-internal"
repeat_region        complement(19330. .19690)
                     /rpt_family="THE1C"
repeat_region        complement(19913. .20102)
                     /rpt_family="MIR"
repeat_region        complement(20317. .20464)
                     /rpt_family="MIR"
repeat_region        complement(20885. .21064)
                     /rpt_family="MIR"
repeat_region        complement(22148. .22499)
                     /rpt_family="MER102"
repeat_region        complement(23923. .24039)
                     /rpt_family="L2"
repeat_region        24227. .24245
                     /rpt_family="AT_rich"
repeat_region        24246. .24286
                     /rpt_family="(TG)n"
repeat_region        complement(24642. .24706)
                     /rpt_family="L2"
repeat_region        26447. .26496
                     /rpt_family="CT-rich"
repeat_region        26833. .26958
                     /rpt_family="MIR"
repeat_region        complement(27491. .27765)
                     /rpt_family="L2"
repeat_region        27962. .28048
                     /rpt_family="MIR"
repeat_region        complement(28060. .28370)
                     /rpt_family="AluSx"
repeat_region        28729. .28948
                     /rpt_family="MER58A"
repeat_region        31019. .31105
                     /rpt_family="GA-rich"
repeat_region        32376. .32437
                     /rpt_family="(TTTTC)n"
repeat_region        33045. .33198
                     /rpt_family="MIR"
repeat_region        33605. .33658
                     /rpt_family="L2"
repeat_region        complement(34014. .34065)
                     /rpt_family="MLT1K"
repeat_region        complement(34240. .34650)
                     /rpt_family="MLT1B"
repeat_region        34737. .35048
                     /rpt_family="AluSq"
repeat_region        36062. .36540
                     /rpt_family="L2"
repeat_region        36903. .37043
                     /rpt_family="MIR"
repeat_region        38116. .38208
                     /rpt_family="MIR"
repeat_region        38963. .39040
                     /rpt_family="L2"
repeat_region        39153. .39328
                     /rpt_family="(TA)n"
unsure               39232. .39285
                     /note="Single-stranded coverage."
repeat_region        39464. .39567
                     /rpt_family="MIR"
repeat_region        complement(39607. .39813)
                     /rpt_family="MLT1I"
repeat_region        39971. .40256
                     /rpt_family="AluSx"
repeat_region        40626. .40672
                     /rpt_family="T-rich"
repeat_region        complement(40792. .41195)
                     /rpt_family="L2"
repeat_region        41242. .41263
                     /rpt_family="AT_rich"
repeat_region        complement(41264. .41551)
                     /rpt_family="AluJo"
repeat_region        41552. .41566
                     /rpt_family="AT_rich"
repeat_region        42464. .42752
                     /rpt_family="AluSx"
repeat_region        42758. .42841
                     /rpt_family="(TA)n"
repeat_region        complement(44068. .44161)
                     /rpt_family="MIR"
repeat_region        45162. .45541
                     /rpt_family="L2"
repeat_region        complement(45549. .45744)
                     /rpt_family="AluJo"
repeat_region        45750. .45847
                     /rpt_family="L2"
repeat_region        47265. .47296
                     /rpt_family="(TTTTG)n"
repeat_region        49777. .49824

alignment_scores:
            Quality:  100.50              Length:    248
              Ratio:  0.791                 Gaps:     13
 Percent Similarity:  51.210   Percent Identity: 24.194

alignment_block:
 US-09-674-779-2 x AC006925   ..

 Align seg 1/1  to: AC006925  from: 1  to: 176186

      18 ValAlaCysSerAlaProIleProThr.....AsnProGlnVa 30
         :::      |||   |||:::||||||           |||   |||::
   24893 ATTATCATCTCCCCTCCAGTTCCCACCAGACTTCTAGGAAACAGCCAGAC 24942

      30 lSerProIleLysThrProSerValLeuIleThrLysAspLysIleGlyA 47
         ::::|||      :::|||   :::   |||  .    :::|||   :::
   24943 AGCTCCTCCAACCAATCCAGTAGCCAAAATAATCCATTCCAAAGCTTCAA 24992

      47 spHisHisThrHisGlu.........HisAspGluSerValSerHisVal 60
           :::|||   :::|||    .      :::      |||::::::|||
   24993 CTGAGCATCTTAATGAAATAATTGTTTACCTGATTAGTATTCAGCATCCG 25042

      61 GlyLeuGlnAlaHis.........PheGluThrTrpLeu...GlnMetHi 73
            :::|||             |||:::   :::|||   |||   ||
   25043 CAGAAGGAGGCTAAAACCCACCCCTTTCAACCTTTCCTCTCTCAGTCTCA 25092

      73 sHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerA 90
         |            ::::::   |||   :::|||   |||::::::
   25093 CTCTCTCTTTCCTTCCCACCCTGTCTCCTTTCAGATATACTTTCGA.... 25138

      90 rgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArg 106
                    :::||||||:::|||                ::::::
   25139 ..............ATCCCACCCCTTTCAGGGGCACAAAGACCAACTCAA 25174

     107 SerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLe 123
            :::         ::::::   |||:::||||||       |||||
   25175 ATCTGATCGCAGCATTATCAATCAACTTATTCCCTCCCCACCGGGCACCT 25224

     123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAsp..... 136
         |         :::   |||   |||   ::::::::::::
   25225 CCTAAGATATTTTTCACCATTTCTTTCATTCTTTGAAAACGATTTCTTTG 25274

     137 ..LeuLysSer.ArgGlyIleLeuProAlaAsnThrGlnIleArgSerVa 152
            |||       |||      ||||||:::             :::::
   25275 TCACCAAGCCTGTGTGGCACTTCCCCTGCCTCTCCTCTTGACCCTGCACT 25324
```

```
    152 lTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMet.... 167
        :   :::   :::   :::|||:::::::|||   |||||||||
  25325 GTAATCTGGACACTCCTGGAGCCAAGCCGTGGGAGAGGCAGCCATGCTCA 25374

    168 ...................SerLysHisLeuThrAsnSerAlaIleAspIle 178
                          :::      |||   |||:::::::::
  25375 GTCTCCCTGGATAAAGCCAGGGTTCCCTTGGCAACGGGACCGTGTGGACA 25424

    179 TrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAs 195
        ::::::      :::            |||:::|||:::::::|||:::::
  25425 TTCGCAAAGGGGTTTCCCTCCACCCCCCAGGTCCTCTTTCAACTCAAGAG 25474

    195 nArg...LeuCysGlnTyrTrp......LeuGluHisGly...GluAsnG 208
        :      ||||||      |||      |||      |||   ||||||:
  25475 CTCAGTCCTGTGTCTCTCATGGAGGCGTCTCTAACCAGGGTAGGAAAACA 25524

    208 lnAsn.PheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHisLeuAspTh 224
        ::::: ::::::::::|||  ::::::      ||||||  |||   ::
  25525 ATAGTTCTAAATATCGGTAATTGGTGTTTGGAGGCCATTGTTTTGCATTC 25574

    224 rGlnGlyPheArg.........LysTrpGlyAlaGlnPhe 234
        :  |||   :::         :::|||::::::::::|||
  25575 AACAGGCAGGAAAATGTGAGTTGAGTGGGCGAGTGAATTC 25614

seq_name: gb_htg:AC106530

seq_documentation_block:
LOCUS       AC106530              194508 bp    DNA     linear   HTG 12-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-50L12, WORKING DRAFT SEQUENCE, 59
            unordered pieces.
ACCESSION   AC106530
VERSION     AC106530.1  GI:18139054
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 194508)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194508)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: GLGG
              Center clone name: CH230-50L12
            -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329First call to
            findPhrapList
              Consensus quality: 169488 bases at least Q40
              Consensus quality: 176660 bases at least Q30
              Consensus quality: 183878 bases at least Q20
              Estimated insert size: 174717; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 59 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11545: contig of 11545 bp in length
            *    11546    11645: gap of unknown length
            *    11646    19125: contig of 7480 bp in length
            *    19126    19225: gap of unknown length
            *    19226    27540: contig of 8315 bp in length
            *    27541    27640: gap of unknown length
            *    27641    34907: contig of 7267 bp in length
            *    34908    35007: gap of unknown length
            *    35008    40495: contig of 5488 bp in length
            *    40496    40595: gap of unknown length
            *    40596    45262: contig of 4667 bp in length
            *    45263    45362: gap of unknown length
            *    45363    51446: contig of 6084 bp in length
            *    51447    51546: gap of unknown length
            *    51547    56966: contig of 5420 bp in length
            *    56967    57066: gap of unknown length
            *    57067    61760: contig of 4694 bp in length
            *    61761    61860: gap of unknown length
            *    61861    66676: contig of 4816 bp in length
            *    66677    66776: gap of unknown length
            *    66777    70702: contig of 3926 bp in length
            *    70703    70802: gap of unknown length
            *    70803    75586: contig of 4784 bp in length
            *    75587    75686: gap of unknown length
            *    75687    81828: contig of 6142 bp in length
            *    81829    81928: gap of unknown length
            *    81929    86907: contig of 4979 bp in length
            *    86908    87007: gap of unknown length
            *    87008    90055: contig of 3048 bp in length
            *    90056    90155: gap of unknown length
            *    90156    95728: contig of 5573 bp in length
```

```
*    95729    95828: gap of unknown length
*    95829    99337: contig of 3509 bp in length
*    99338    99437: gap of unknown length
*    99438   103719: contig of 4282 bp in length
*   103720   103819: gap of unknown length
*   103820   106877: contig of 3058 bp in length
*   106878   106977: gap of unknown length
*   106978   110433: contig of 3456 bp in length
*   110434   110533: gap of unknown length
*   110534   113415: contig of 2882 bp in length
*   113416   113515: gap of unknown length
*   113516   117232: contig of 3717 bp in length
*   117233   117332: gap of unknown length
*   117333   120913: contig of 3581 bp in length
*   120914   121013: gap of unknown length
*   121014   124626: contig of 3613 bp in length
*   124627   124726: gap of unknown length
*   124727   128540: contig of 3814 bp in length
*   128541   128640: gap of unknown length
*   128641   132026: contig of 3386 bp in length
*   132027   132126: gap of unknown length
*   132127   134301: contig of 2175 bp in length
*   134302   134401: gap of unknown length
*   134402   137766: contig of 3365 bp in length
*   137767   137866: gap of unknown length
*   137867   139560: contig of 1694 bp in length
*   139561   139660: gap of unknown length
*   139661   142247: contig of 2587 bp in length
*   142248   142347: gap of unknown length
*   142348   144802: contig of 2455 bp in length
*   144803   144902: gap of unknown length
*   144903   146978: contig of 2076 bp in length
*   146979   147078: gap of unknown length
*   147079   148542: contig of 1464 bp in length
*   148543   148642: gap of unknown length
*   148643   151409: contig of 2767 bp in length
*   151410   151509: gap of unknown length
*   151510   154038: contig of 2529 bp in length
*   154039   154138: gap of unknown length
*   154139   155704: contig of 1566 bp in length
*   155705   155804: gap of unknown length
*   155805   158058: contig of 2254 bp in length
*   158059   158158: gap of unknown length
*   158159   160106: contig of 1948 bp in length
*   160107   160206: gap of unknown length
*   160207   162081: contig of 1875 bp in length
*   162082   162181: gap of unknown length
*   162182   163696: contig of 1515 bp in length
*   163697   163796: gap of unknown length
*   163797   165744: contig of 1948 bp in length
*   165745   165844: gap of unknown length
*   165845   167567: contig of 1723 bp in length
*   167568   167667: gap of unknown length
*   167668   169690: contig of 2023 bp in length
*   169691   169790: gap of unknown length
*   169791   171051: contig of 1261 bp in length
*   171052   171151: gap of unknown length
*   171152   173106: contig of 1955 bp in length
*   173107   173206: gap of unknown length
*   173207   174506: contig of 1300 bp in length
*   174507   174606: gap of unknown length
*   174607   176921: contig of 2315 bp in length
*   176922   177021: gap of unknown length
*   177022   179060: contig of 2039 bp in length
*   179061   179160: gap of unknown length
*   179161   180838: contig of 1678 bp in length
*   180839   180938: gap of unknown length
*   180939   182582: contig of 1644 bp in length
*   182583   182682: gap of unknown length
*   182683   184049: contig of 1367 bp in length
*   184050   184149: gap of unknown length
*   184150   185764: contig of 1615 bp in length
*   185765   185864: gap of unknown length
*   185865   187483: contig of 1619 bp in length
*   187484   187583: gap of unknown length
*   187584   188623: contig of 1040 bp in length
*   188624   188723: gap of unknown length
*   188724   189759: contig of 1036 bp in length
*   189760   189859: gap of unknown length
*   189860   190958: contig of 1099 bp in length
*   190959   191058: gap of unknown length
*   191059   192233: contig of 1175 bp in length

alignment_scores:
           Quality:  100.50              Length:    174
             Ratio:  1.415                 Gaps:      6
 Percent Similarity:  40.805   Percent Identity: 22.414

alignment_block:
 US-09-674-779-2 x AC106530/rev   ..

 Align seg 1/1  to reverse of: AC106530  from: 1  to: 194508

       4 PheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuValAlaCy 20
         |||:::   ||||||          |||:::|||:::   |||:::||
   39372 TTTACTTTTTATTTTTCTAAACTTTCTATTGATTCTTTGTGAGTCTCATG 39323

      20 sSerAlaProIleProThrAsnProGlnValSerProIleLysThrProS 37
         |   :::           :::|||           |||:::|||
   39322 TCATACACACCAGTCTCACTCACCTCCCCATCCCCTCATATCTACCCNTT 39273

      37 erValLeuIle...ThrLysAspLysIleGlyAspHis...HisThrHis 51
           |||   :::   |||      :::         |||   |||||||||
   39272 TTGTGTGTGTTCACACACACACACGCACACGCACACATGCACACACACAC 39223

      52 GluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluTh 68
           |||   :::      :::|||:::      ::::::|||::::::||
   39222 ACGCACACACACATGTACACGCACACACACACGCACACACACATGCACAC 39173

      68 rTrpLeuGlnMetHis.....................HisAlaThrLysG 78
         |  ::::::   |||                      |||:::      :
   39172 ACACATGCACACACACATGTATACACACACACACACACACACACACAC 39123

      78 lnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr 94
         ::         |||      :::|||      |||   ::::::::::::
   39122 ACACACACCATAGAAAACCATCTTATTGTGGAAGCTGTATGTCACAGCAT 39073

      95 LeuPro.........ProMetSerGln....................... 100
         :::|||         |||:::|||
   39072 GTCCCACAGTGTATCCCTCTGTCCACACATCTTCACTTGAGAATGTTCAC 39023

     100 ..................................................... 100

   39022 TGCAGTAAGTCACTGGTCTGGTTTGAGGTCTTGGCTTCCATGACACCATC 38973

     101 ........LeuLeuThrThrAlaArgSerTrpGlnAlaCys......... 111
                 |||||||||   :::   ||||||   :::|||
   38972 AACACTGGCTCCTCACTGGGACTCCCTCCTGGTTATCCTGTTGTTGTGCT 38923

     112 ................................GlyHisGluProTyrGl 117
                                        |||   :::|||   ::
   38922 GTGTCATGGAGGTCCTGCAGCTTTGAATCTGCAGGACCAGACCCTTCACA 38873

     117 nLeuProProGluHisLeuTrp 124
         ::::|||      |||   |||
   38872 TGTCCCAGCAGTTCACAGATGG 38851

seq_name: gb_in:CAE223616

seq_documentation_block:
LOCUS       CAE223616              4383 bp    mRNA    linear   INV 24-FEB-1999
DEFINITION  Caehorhabditis elegans mRNA for calcium ATPase.
ACCESSION   AJ223616
VERSION     AJ223616.1  GI:2826863
KEYWORDS    calcium ATPase; mca-1 gene.
SOURCE      Caenorhabditis elegans.
```

```
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 4383)
  AUTHORS   Kraev,A.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-1998) Kraev A.S., Laboratory of Biochemistry III,
            Swiss Federal Institute of Technology, Universitaetsstr. 16,
            Zurich, CH-8092, SWITZERLAND
REFERENCE   2  (bases 1 to 4383)
  AUTHORS   Kraev,A., Kraev,N. and Carafoli,E.
  TITLE     Identification and functional expression of the plasma membrane
            calcium ATPase gene family from Caenorhabditis elegans
  JOURNAL   J. Biol. Chem. 274 (7), 4254-4258 (1999)
  MEDLINE   99134356
FEATURES             Location/Qualifiers
     source          1. .4383
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /chromosome="IV"
                     /clone="yk49f8 (765-4383)"
                     /clone_lib="Yuji Kohara cDNA"
                     /dev_stage="mixed"
     misc_feature    1. .22
                     /function="trans-spliced leader SL1"
                     /evidence=experimental
     gene            84. .3770
                     /gene="mca-1"
     CDS             84. .3770
                     /gene="mca-1"
                     /codon_start=1
                     /product="calcium ATPase"
                     /protein_id="CAA11491.1"
                     /db_xref="GI:2826864"
                     /db_xref="SPTREMBL:O45215"
                     /translation="MRLRRGHTYDEDAKEFGCSLGDLRGLMEARGAEAIVRLSTEHEG
                     VEGLCKKLKTDSLVGLNGEQADLDRRRHVYGANTIPPAKSKGFVRLVLDACKDPTLVI
                     LVLSGFINLALSFYEPTSAAEDATQHLVNATTAAILANGTFMSTTEAPSEGHGTAWIE
                     GVAILLCVIVVVLVTAVNDYSKERQFRSLQEKIETGQKFSVIRNGEAIDVPVSDLVVG
                     DIARVKYGDLLPADGFLIQSNDLKIDESSLTGESDHIKKSIESDPVLLSGTYAMEGSG
                     KMLITAVGVNSQTGIIMTLLGAGKAGIGDDDSTSTSSSSSSSSSSGSSSNGSSDSSK
                     SGDDDLTAKSVLQAKLSKLALQIIYCGTTIAIIALIVLVTRFCLDHYVFEKNEFSLVD
                     IQMFVKFFIIAVTILVISIPEGLPLAIALALTYSVRKMMHDNNLVRHLDACETMGNAT
                     SICSDKTGTLTTNRMTVVQSYINGNHYTSQEAQPHGANLPGSTGPILMEAISVNCAYN
                     SMIVEPTKAGEQIQQLGNKTECGLLGFVNRLGGDYAAIRKKFPEHDLTKVYTFNSSRK
                     CMMTVVPYAENGQNIGYRVYCKGASEIVLGRCTYLIGSDGKPHQLTGDRLKEITSTII
                     HEMANSGLRTICVAYKTIIKKGTRDVEKTEIEFAEDSDIDWDDEDAMYQNFTGIAICG
                     IQDPVRPEVPVAISKCKKAGITVRMVTGDNIMTARAIAMSCKILEPGEDFLALEGKEF
                     NERIRDENGKVSQAKLDEIWPRLRVLARAQPADKYTLVKGIIDSKATPQREIVAVTGD
                     GTNDGPALKKADVGFAMGIAGTDVAKEASDIILTDDNFTSIVKAVMWGRNVYDSISKF
                     LQFQLTVNVVAVITAFVGAVTVSDSPLKAVHMLWINLIMDTLASLALATEQPTDELLE
                     RKPYGRKKSLISRTMVKNILCHALYQLIIIFVIFFYGDTIFGIKTGLYAPLFAPPSQH
                     FTLVFNAFVMMTVFNEINARKVHGERNVFKGLASNRVFCVIWVTTFIAQIIIVQFGGA
                     WFSTAPLTLQQWIVCLVLGFSTLIWGQIVATIPSKKLPKAWKVGKGEVQPANLHINGD
                     YNVRARSRAVTLRRSGKSLWVRGMFIIGNHLRVLRAFGMEKSEKAAFGRTAPAMTAEA
                     AERWRASYRKYRHQKHQEKKATAETAESVKSADWAKEQKEKKKTFKQIKQVARGKSLD
                     KDSKKHHKKRKDQTNVDMEDIELN"
BASE COUNT     1248 a    975 c    925 g   1235 t
ORIGIN


alignment_scores:
            Quality:  100.00            Length:    264
              Ratio:  0.781               Gaps:     11
 Percent Similarity:  48.485   Percent Identity: 21.212

alignment_block:
 US-09-674-779-2 x CAE223616   ..

 Align seg 1/1  to: CAE223616  from: 1  to: 4383

       2 LysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuVa 18
         ::::::::::::::::   |||::::::   |||::::::::::::   ::
      2863 GAACAATGGTCAAGAATATTCTTTGTCATGCTCTTTACCAACTTATCATT 2912

        18 lAlaCysSerAlaProIleProThrAsnProGlnValSerProIleLysT 35
           ::::   |||:::                              |||      :::|
      2913 ATCTTTGTCATCT.......................TCTTCTACGGAGA 2938

        35 hrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHis 51
           |||||||:::         :::|||::::::     ||||||:::
      2939 CACCATCTTTGGAATCAAGACAGGACTCTATG......CACCACTCTTCG 2982

        52 GluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluTh 68
             |||   :::::::::|||
      2983 CTCCACCATCACAACATTTCA.........................CT 3005

        68 rTrpLeuGlnMetHis..............HisAlaThrLysGlnGluV 80
           |||::::::||||||                :::   ::::::   :
      3006 TTGGTTTTCAATGCATTCGTTATGATGACTGTCTTCAATGAGATCAATGC 3055

        80 alValArgTyr..................................... 83
           ::||||||:::
      3056 TCGTAAGGTTCACGGAGAGCGTAACGTGTTTAAAGGATTGGCATCTAACC 3105

        84 ........................GlnAlaTyrLeuGlnSerArgLe 91
                                   ::::::   |||:::|||   ||
      3106 GAGTTTTCTGTGTCATTTGGGTTACAACATTCATTGCTCAAATCATTATT 3155

        91 uGlyAsnTyrLeuProProMetSerGlnLeuLeu..ThrThrAlaArgSe 107
           |   |||         |||   ||||||||||     |||   :::::
      3156 GTTCAATTCGGAGGTGCCTGGTTCTCAACTGCTCCACTTACGCTTCAACA 3205

       107 rTrpGlnAlaCys.........GlyHisGluProTyrGlnLeuProProG 121
           :|||   :::|||         |||   :::
      3206 ATGGATTGTATGCCTCGTTCTTGGATTCTCAACTCTT.............. 3242

       121 luHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu 137
              :::|||||||||||||||   |||:::        :::
      3243 .....ATCTGGGGACAAATTGTGGCAACTATCCCATCAAAGAAGCTACCA 3287

       138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
           |||:::      :::      ||||||||||      |||      :::::::::
      3288 AAGGCCTGGAAAGTTGGAAAGGGAGAAGTTCAACCAGCCAATCTTCACAT 3337

       154 gAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
           |||   :::   |||      |||  :::   |||::::::      |
      3338 CAATGGAGACTACAATGTCCGTGCTCGTTCCCGTGCGGTTACT......C 3381

       171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
           ||   :::|||:::   ::::::||||||      :::   |||   :::
      3382 TTCGCAGATCTGGAAAGTCACTCTGGGTTAGAGGAATGTTCATTATTGGA 3431

       188 GlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHi 204
           :::   |||      |||:::                :::   :::|||
      3432 AACCATCTCCGCGTTCTTCGT............GCTTTCGGAATGGAAAA 3469

       204 sGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleH 221
           :::|||:::      ||||||                        |||
      3470 GAGTGAAAAGGCTGCCTTCGGA..............CGTACTGCTCCAG 3504

       221 isLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnPhe 234
             :::   :::::::::   :::::::|||   |||::::::
      3505 CAATGACGGCTGAAGCTGCTGAACGCTGGCGTGCCAGTTAC 3545

seq_name: gb_pr:HS833B7

seq_documentation_block:
LOCUS       HS833B7                 86574 bp    DNA     linear   PRI 12-DEC-1999
DEFINITION  Human DNA sequence from clone CTA-833B7 on chromosome 22q12.3-13.2
            Contains the NCF4 gene for cytosolic neutrophil factor 4 (40kD),
            the 5' part of the CSF2RB gene for granulocyte-macrophage
            low-affinity colony stimulating factor 2 receptor beta, ESTs, STSs
            and GSSs, complete sequence.
ACCESSION   AL008637
```

```
VERSION     AL008637.1  GI:3136000
KEYWORDS    HTG; colony stimulating factor; CSF2RB; cytosolic neutrophil
            factor; NCF4.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86574)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 18, 1998 this sequence version replaced gi:2578146.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-833B7 is
            from the human BAC library described in U-J. Kim et al. (1996)
            Genomics 34, 213-218.
            VECTOR: pBeloBAC11
            IMPORTANT: This sequence is not the entire insert of clone
            CTA-833B7 It may be shorter because we only sequence overlapping
            sections once, or longer because we arrange for a small overlap
            between neighbouring submissions.
            The true right end of clone CITF22-24E5 is at 100 in this sequence.
            The start of this sequence overlaps with sequence Z82185.
FEATURES             Location/Qualifiers
     source          1. .86574
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q12.3-13.2"
                     /clone="CTA-833B7"
                     /clone_lib="CIT978SK-A2"
     repeat_region   1. .176
                     /note="AluSp repeat: matches 138. .313 of consensus"
     repeat_region   862. .903
                     /note="MIR repeat: matches 212. .254 of consensus"
     repeat_region   1791. .2095
                     /note="AluJb repeat: matches 1. .307 of consensus"
     misc_feature    complement(2600. .3449)
                     /note="match: GSS: Em:AQ749444"
     repeat_region   3359. .3522
                     /note="MER58A repeat: matches 43. .207 of consensus"
     repeat_region   3532. .3838
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   4030. .4288
                     /note="MLT1J repeat: matches 1. .261 of consensus"
     misc_feature    complement(4402. .4679)
                     /note="match: GSS: Em:AQ321484"
     repeat_region   5050. .5195
                     /note="LTR41 repeat: matches 4. .144 of consensus"
     repeat_region   5291. .5370
                     /note="L1ME1 repeat: matches 5526. .5607 of consensus"
     repeat_region   5371. .5676
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   5677. .6252
                     /note="L1ME1 repeat: matches 5607. .6163 of consensus"
     repeat_region   6302. .6485
                     /note="L2 repeat: matches 2078. .2265 of consensus"
     repeat_region   6873. .7196
                     /note="54 copies 6 mer gtgtgt 66 conserved"
     repeat_region   6874. .7197
                     /note="27 copies 12 mer 66 conserved"
     repeat_region   6879. .7190
                     /note="13 copies 24 mer 67 conserved"
     repeat_region   6884. .7195
                     /note="156 copies 2 mer tg 67 conserved"
     repeat_region   6886. .7193
                     /note="22 copies 14 mer 67 conserved"
     repeat_region   6907. .7194
                     /note="9 copies 32 mer 69 conserved"
     repeat_region   7203. .7332
                     /note="L2 repeat: matches 2602. .2750 of consensus"
     repeat_region   7347. .7532
                     /note="MIR repeat: matches 2. .191 of consensus"
     repeat_region   7537. .7640
                     /note="MIR repeat: matches 158. .262 of consensus"
     repeat_region   8152. .8228
                     /note="MIR repeat: matches 184. .260 of consensus"
     repeat_region   8253. .8341
                     /note="L2 repeat: matches 2651. .2748 of consensus"
     repeat_region   8374. .8479
                     /note="MIR repeat: matches 48. .167 of consensus"
     repeat_region   11386. .11858
                     /note="L2 repeat: matches 2164. .2749 of consensus"
     mRNA            join(11913. .12573,15415. .15499,16289. .16442,18762. .18832,
                     21785. .21912,23022. .23079,23697. .23795,27023. .27153,
                     27399. .27464,28998. .29288)
                     /gene="NCF4"
                     /note="match: cDNAs: Em:AB002665 Em:X77094 Em:U59488;
                     match: ESTs: Em:AA177839 Em:AI007048 Em:AA465462
                     Em:AA465389 Em:AA485518 Em:AI381940 Em:AA744805
                     Em:AA975113 Em:AA969460 Em:AA948430 Em:AI435296
                     Em:AI299103 Em:AA648472 Em:AA702857 Em:AI088359 Em:D20144
                     Em:AA909156 Em:AA688071 Em:W95229 Em:AI439568"
                     /product="bK833B7.1 (neutrophil cytosolic factor 4
                     (40kD))"
                     /evidence=not_experimental
     gene            11913. .29288
                     /gene="NCF4"
     repeat_region   11982. .12109
                     /note="L2 repeat: matches 2356. .2502 of consensus"
     misc_feature    join(12483. .12573,15415. .15499,16289. .16442,18762. .18832,
                     21785. .21912,23022. .23079,23697. .23795,27023. .27464,
                     28998. .29288)
                     /gene="NCF4"
                     /note="match: STS: Em:X77094"
     CDS             join(12542. .12573,15415. .15499,16289. .16442,18762. .18832,
                     21785. .21912,23022. .23079,23697. .23795,27023. .27153,
                     27399. .27464,28998. .29193)
                     /gene="NCF4"
                     /note="match: proteins: Tr:O60808 Tr:P97369"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bK833B7.1 (neutrophil cytosolic factor 4
                     (40kD))"
                     /protein_id="CAA15486.1"
                     /db_xref="GI:3136001"
                     /db_xref="SPTREMBL:O60808"
                     /translation="MAVAQQLRAESDFEQLPDDVAISANIADIEEKRGFTSHFVFVIE
                     VKTKGGSKYLIYRRYRQFHALQSKLEERFGPDSKSSALACTLPTLPAKVYVGVKQEIA
                     EMRIPALNAYMKSLLSLPVWVLMDEDVRIFFYQSPYDSEQVPQALRRLRPRTRKVKSV
                     SPQGNSVDRMAAPRAEALFDFTGNSKLELNFKAGDVIFLLSRINKDWLEGTVRGATGI
                     FPLSFVKILKDFPEEDDPTNWLRCYYYEDTISTIKDIAVEEDLSSTPLLKDLLELTRR
                     EFQREDIALNYRDAEGDLVRLLSDEDVALMVRQARGLPSQKRLFPWKLHITQKDNYRV
                     YNTMP"
     repeat_region   12766. .12882
```

```
                     /note="MIR repeat: matches 20. .136 of consensus"
     repeat_region   12944. .13247
                     /note="L2 repeat: matches 2425. .2750 of consensus"
     repeat_region   13315. .13403
                     /note="MIR repeat: matches 53. .143 of consensus"
     repeat_region   14202. .14242
                     /note="L2 repeat: matches 2701. .2741 of consensus"
     repeat_region   14289. .14336
                     /note="2 copies 24 mer 95 conserved"
     repeat_region   14545. .14625
                     /note="L2 repeat: matches 2637. .2728 of consensus"
     repeat_region   14656. .14743
                     /note="MIR repeat: matches 89. .179 of consensus"
     repeat_region   14842. .14977
                     /note="L2 repeat: matches 2213. .2350 of consensus"
     repeat_region   14983. .15050
                     /note="L2 repeat: matches 2638. .2710 of consensus"
     repeat_region   15130. .15225
                     /note="3 copies 32 mer 95 conserved"
     repeat_region   15616. .15815
                     /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region   16065. .16102
                     /note="19 copies 2 mer tt 81 conserved"
     repeat_region   16171. .16218
                     /note="16 copies 3 mer cct 77 conserved"
     repeat_region   16173. .16214
                     /note="7 copies 6 mer tcctcc 81 conserved"
     repeat_region   17561. .17987
                     /note="L2 repeat: matches 1251. .1706 of consensus"
     repeat_region   18193. .18242
                     /note="L2 repeat: matches 2702. .2750 of consensus"
     repeat_region   18243. .18412
                     /note="MER5B repeat: matches 1. .175 of consensus"
     repeat_region   18413. .18473
                     /note="L2 repeat: matches 2638. .2702 of consensus"
     misc_feature    complement(18442. .18745)
                     /note="match: STS: Em:HSA046ZA5; match: STS: Em:Z51609"
     repeat_region   18481. .18516
                     /note="3 copies 12 mer 94 conserved"
     misc_feature    complement(18698. .19137)
                     /note="match: GSS: Em:AQ499165"
     repeat_region   19252. .19406
                     /note="MER91A repeat: matches 11. .193 of consensus"
     repeat_region   19642. .19730
                     /note="MIR repeat: matches 57. .145 of consensus"
     repeat_region   20478. .20578
                     /note="MIR repeat: matches 9. .104 of consensus"

alignment_scores:
            Quality:  100.00              Length:    130
              Ratio:  1.493                 Gaps:      6
 Percent Similarity:  51.538   Percent Identity: 30.000

alignment_block:
 US-09-674-779-2 x HS833B7/rev   ..

 Align seg 1/1  to reverse of: HS833B7  from: 1  to: 86574

       9 IleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIlePr 25
         :::|||:::::::::::::|||   |||:::|||      :::||||||
   24268 CTCACCTCCATCCTCACCTCCAACCTTACTGCCAGTCTCACCCCCATCCT 24219

      25 oThrAsnProGlnValSerProIleLysThrProSerValLeuIleThrL 42
         ||||::      :::|||   |||   |||   :::::::::   |||
   24218 CACGTCCACCCTTACCTCC...ATCCTCACCTTCACCCTCATCTCCACC. 24173

      42 ysAspLysIleGlyAspHis...HisThrHisGluHisAspGluSerVal 57
               ::::::   |||   |||   |||   |||   :::
   24172 ........CTCAATCTCCATCCTCACCTTCACCCTCATCTCCACCCTCAC 24131

      58 SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHis.. 73
         |||      ||||||:::|||:::::::::   ||||||   |||
   24130 CTCCATCCTCACCTCCAATCTCATCTCCAATCTCACCTCCAACCTCACCT 24081

      74 .......HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuG 88
               |||                        :::   :::   :::||||
   24080 CCAATCTCACCTCCAACCTTACCTCCAATCTCACCTCCATCCTCACCTCC 24031

      88 lnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThr 104
         |||||:::|||   :::||||||   |||
   24030 AATCTCATCTCCACTCTTACCTCCACCCT..................... 24002

     105 AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGl 121
                                 |||         :::|||   |||
   24001 .....................CACCTCCAATCTCACCTTCATCCT.. 23978

     121 uHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyr 134
         ||||||   :::::::::   |||   ||||||   |||
   23977 .CACCTCCAATCTCACCTCCATCCTCACCTCCACTCTTAC 23939

seq_name: gb_pr:AC083964

seq_documentation_block:
LOCUS       AC083964              161160 bp    DNA     linear   PRI 25-NOV-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-91J19, complete sequence.
ACCESSION   AC083964
VERSION     AC083964.3  GI:15290804
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161160)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-91J19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161160)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 161160)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
```

```
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 26, 2001 this sequence version replaced gi:13249446.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11276
                Center clone name: 91_J_19
            --------------.
FEATURES             Location/Qualifiers
     source          1. .161160
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-91J19"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   269. .478
                     /rpt_family="MIR"
     repeat_region   complement(515. .619)
                     /rpt_family="LTR67"
     repeat_region   complement(762. .877)
                     /rpt_family="LTR67"
     repeat_region   922. .1112
                     /rpt_family="L2"
     repeat_region   complement(2229. .2408)
                     /rpt_family="AluSq"
     repeat_region   complement(2409. .8561)
                     /rpt_family="L1PA4"
     repeat_region   8650. .8677
                     /rpt_family="AT_rich"
     repeat_region   complement(9721. .10258)
                     /rpt_family="MER44D"
     repeat_region   10397. .10592
                     /rpt_family="(CAGG)n"
     repeat_region   11100. .11293
                     /rpt_family="(CAGG)n"
     repeat_region   complement(11661. .11878)
                     /rpt_family="MER44D"
     repeat_region   complement(12853. .12980)
                     /rpt_family="Charlie8"
     repeat_region   16234. .16386
                     /rpt_family="MLT1I"
     repeat_region   18474. .18642
                     /rpt_family="(CAGAGA)n"
     repeat_region   18648. .18716
                     /rpt_family="GA-rich"
     repeat_region   20232. .20266
                     /rpt_family="AT_rich"
     repeat_region   complement(20267. .20376)
                     /rpt_family="FLAM_A"
     repeat_region   complement(22281. .22591)
                     /rpt_family="AluJb"
     repeat_region   22670. .22981
                     /rpt_family="AluY"
     repeat_region   23415. .23524
                     /rpt_family="HAL1b"
     repeat_region   23608. .23900
                     /rpt_family="AluSp"
     repeat_region   23902. .23986
                     /rpt_family="HAL1b"
     repeat_region   complement(24615. .24922)
                     /rpt_family="AluSc"
     repeat_region   complement(25637. .25843)
                     /rpt_family="L1MC5"
     repeat_region   complement(26019. .26341)
                     /rpt_family="L1"
     repeat_region   26342. .26362
                     /rpt_family="(TTTG)n"
     repeat_region   complement(26363. .26450)
                     /rpt_family="L1"
     repeat_region   28942. .29051
                     /rpt_family="L2"
     repeat_region   29319. .29394
                     /rpt_family="MIR"
     repeat_region   complement(29537. .29684)
                     /rpt_family="MER104"
     repeat_region   31106. .31413
                     /rpt_family="L1MCa"
     repeat_region   31414. .31716
                     /rpt_family="AluSx"
     repeat_region   complement(32355. .32490)
                     /rpt_family="MER5B"
     repeat_region   32634. .32718
                     /rpt_family="L1ME4A"
     repeat_region   32798. .32835
                     /rpt_family="AT_rich"
     repeat_region   complement(33424. .33711)
                     /rpt_family="L2"
     repeat_region   complement(36020. .36178)
                     /rpt_family="MER63A"
     repeat_region   36258. .36438
                     /rpt_family="MLT1A1"
     repeat_region   37084. .37743
                     /rpt_family="L1PA13"
     repeat_region   complement(37744. .37837)
                     /rpt_family="AluSc"
     repeat_region   37839. .38276
                     /rpt_family="L1PA13"
     repeat_region   38277. .38347
                     /rpt_family="L1PA13"
     repeat_region   38348. .38418
                     /rpt_family="L1PA13"
     repeat_region   38419. .38489
                     /rpt_family="L1PA13"
     repeat_region   38490. .38559
                     /rpt_family="L1PA13"
     repeat_region   38560. .38629
                     /rpt_family="L1PA13"
     repeat_region   38630. .38699
                     /rpt_family="L1PA13"
     repeat_region   38700. .38805
                     /rpt_family="L1PA13"
     repeat_region   complement(38979. .39045)
                     /rpt_family="MIR"
     repeat_region   41036. .41093
                     /rpt_family="MER4A"
     repeat_region   41104. .41146
                     /rpt_family="(CA)n"
     repeat_region   41150. .41489
                     /rpt_family="MER4C"
     repeat_region   41676. .41739
                     /rpt_family="GA-rich"
     repeat_region   complement(41963. .42083)
                     /rpt_family="MIR"
     repeat_region   44242. .44476
                     /rpt_family="MIR"
     repeat_region   47025. .47406
                     /rpt_family="L2"
     repeat_region   complement(47406. .47660)
                     /rpt_family="L2"
     repeat_region   complement(47714. .47870)
```

```
                    /rpt_family="MER5B"

alignment_scores:
            Quality:  100.00              Length:    201
              Ratio:  1.136                 Gaps:      9
 Percent Similarity:  43.781    Percent Identity: 19.900

alignment_block:
 US-09-674-779-2 x AC083964   ..

 Align seg 1/1  to: AC083964  from: 1  to: 161160

      23 ProIleProThrAsn..................................... 27
         |||:::|||||||:::
  139277 CCCGTGCCCACCTCAGCACGCACCAACACGGAATCTAGAGCCATGCATCG 139326

      28 .............ProGlnValSerProIleLysThrPro.......... 36
                      ||||||   :::|||   ::::::|||
  139327 GAACACGTGCCCACCTCAGCACACACCAACACGGAATCCAGAGCAGCACA 139376

      37 ................SerValLeuIleThrLysAspLysIleGlyAsp 47
                         |||      |||:::   ::::::   :::
  139377 CTGGAACCAATGCCCACTCACCATGCATCAACACAGAATCCAAAGCCACA 139426

      48 His.....................HisThrHisGluHisAspGluServa 57
         |||                     |||:::   :::   ::::::   ::
  139427 CATCAGAACCTGTACCCACCTCAGCACGCACCCAACACGGAATCCAGAGC 139476

      57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH 74
         ::::|||:::|||   :::||||||::::::|||           :::|
  139477 CACACACACTGGAACCAATGCCCACCTCAACACA.........CACCAAC 139517

      74 isAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg 90
         ||:::        ::::::      :::|||     :::||||||
  139518 ACGGAATCCACGCCCACATCAGCACACACCAGCATGGAATCCAGAGCTAC 139567

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSe 107
            ::::::   :::|||      :::::::::   |||   :::||||::
  139568 ACATCAGAACCCGTGCCCACCTCAGCACGCATCAACACGGAATCCAGAGC 139617

     107 r......TrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluH 122
         :      |||:::  |||                    ||||||:::|
  139618 CACACACTGGAACCCATGC.....................CCACCTCAGC 139646

     122 is......LeuTrpGlyGlnIleValProThr..................... 130
         ||      :::|||:::|||         |||
  139647 ACGCGTCAATATGGAATCAAGAGCCACACACTGGAACCAATGCCCACCTC 139696

     131 .....................................LeuHisLeuTy 134
                                             |||   ::
  139697 TGCACACACCAACACGGAATCCAGAGCCACACAAGGGTAACCCACCCCCA 139746

     134 rGlnAspLeuLysSerArgGlyIle.....................LeuP 144
         :   :::      ::::::||||||                    :::|
  139747 CCTCAGCACGCACCAACATGGAATCCAGAGCCACACAATGGAACCAATGC 139796

     144 roAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGln 160
         ||:::::::::|||::::::         |||      :::   ||||||
  139797 CCACCTCAGCACAAGTCAACATGGAATCAAGAGCCACACACTGGAACCAA 139846

     161 Cys 161
         |||
  139847 TGC 139849

seq_name: gb_htg:AC090135

seq_documentation_block:
LOCUS       AC090135              190173 bp    DNA     linear   HTG 19-JAN-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-18D5 map 8, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.
ACCESSION   AC090135
VERSION     AC090135.9  GI:18250010
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190173)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-18D5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190173)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 19, 2002 this sequence version replaced gi:18182767.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11647
                Center clone name: 18_D_5
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25242: contig of 25242 bp in length
            *    25243 25342: gap of      100 bp
            *    25343   120061: contig of 94719 bp in length
            *   120062 120161: gap of      100 bp
            *   120162   181820: contig of 61659 bp in length
            *   181821 181920: gap of      100 bp
            *   181921   190173: contig of 8253 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190173
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-18D5"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    53015 a  40688 c  41908 g  54244 t    318 others
ORIGIN
```

```
alignment_scores:
            Quality:  100.00              Length:    201
              Ratio:  1.136                 Gaps:      9
 Percent Similarity:  43.781   Percent Identity: 19.900

alignment_block:
 US-09-674-779-2 x AC090135/rev   ..

 Align seg 1/1  to reverse of: AC090135  from: 1  to: 190173

      23 ProIleProThrAsn..................................... 27
         |||:::|||||||:::
   57685 CCCGTGCCCACCTCAGCACGCACCAACACGGAATCTAGAGCCATGCATCG 57636

      28 .............ProGlnValSerProIleLysThrPro.......... 36
                     ||||||   :::|||   ::::::|||
   57635 GAACACGTGCCCACCTCAGCACACACCAACACGGAATCCAGAGCAGCACA 57586

      37 ................SerValLeuIleThrLysAspLysIleGlyAsp 47
                         |||      |||:::   ::::::   :::
   57585 CTGGAACCAATGCCCACTCACCATGCATCAACACAGAATCCAAAGCCACA 57536

      48 His.....................HisThrHisGluHisAspGluSerVa 57
         |||                    |||:::   :::   ::::::   ::
   57535 CATCAGAACCTGTACCCACCTCAGCACGCACCCAACACGGAATCCAGAGC 57486

      57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH 74
         ::::|||:::|||   :::||||||:::::::|||           :::|
   57485 CACACACACTGGAACCAATGCCCACCTCAACACA.........CACCAAC 57445

      74 isAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg 90
         ||:::         ::::::      :::|||      :::||||||
   57444 ACGGAATCCACGCCCACATCAGCACACACCAGCATGGAATCCAGAGCTAC 57395

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSe 107
            ::::::   :::|||      :::::::::   |||   :::|||::
   57394 ACATCAGAACCCGTGCCCACCTCAGCACGCATCAACACGGAATCCAGAGC 57345

     107 r......TrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluH 122
         :      |||:::   |||                  ||||||:::|
   57344 CACACACTGGAACCCATGC.....................CCACCTCAGC 57316

     122 is.....LeuTrpGlyGlnIleValProThr..................... 130
         ||      :::|||:::|||         |||
   57315 ACGCGTCAATATGGAATCAAGAGCCACACACTGGAACCAATGCCCACCTC 57266

     131 .....................................LeuHisLeuTy 134
                                             |||   ::
   57265 TGCACACACCAACACGGAATCCAGAGCCACACAAGGGTAACCCACCCCCA 57216

     134 rGlnAspLeuLysSerArgGlyIle.....................LeuP 144
         :   :::      ::::::||||||                    :::|
   57215 CCTCAGCACGCACCAACATGGAATCCAGAGCCACACAATGGAACCAATGC 57166

     144 roAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGln 160
         ||:::::::::|||::::::      |||      :::   ||||||
   57165 CCACCTCAGCACAAGTCAACATGGAATCAAGAGCCACACACTGGAACCAA 57116

     161 Cys 161
         |||
   57115 TGC 57113

seq_name: gb_ba:AP003008

seq_documentation_block:
LOCUS       AP003008              346820 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 15/21.
ACCESSION   AP003008 BA000012
VERSION     AP003008.2  GI:14025721
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 346820)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994983.
FEATURES             Location/Qualifiers
     source          1. .346820
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
     gene            complement(480. .1706)
                     /gene="mll5963"
     CDS             complement(480. .1706)
                     /gene="mll5963"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52321.1"
                     /db_xref="GI:14025722"
                     /translation="MARRALILVEGHRSNGPLYMQAIQNLGLRPITLSTDPAQYDYLK
                     ARKLEAIRVDTDSIEALIHECSGLRGSYDIAGITGFTALDESLYSRIGKLCQYFSLPG
                     PNPESIEQCCDKFTQRQFLAAAGVPIPVYRLAVNEAEVESAAAEIGLPVILKPAAGSG
                     SSGVRLCRNVDELAEHTTYLLGEKHLWSSPRILVEEFAQGPFYSAEIMGNEVVGIEAT
                     DFGPPPHFVIRGATHPAPLTDEEYERIADVSLSCLPALGLGWGPANIELRWTKHGPVI
                     IEVNPRLSGATCPKLVRLAYGVDLVAEHIKLVIGDECDLRRGHSNAAATRFLLPDADG
                     ILDCIDGESRAAAVPGVAEVKFYVGPKTPIVRKGDYRDPIGHVIAASPTRAQTEAILQ
                     DAVDMLAWSITPSRDE"
     gene            complement(2451. .3617)
                     /gene="mll5964"
     CDS             complement(2451. .3617)
                     /gene="mll5964"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52322.1"
                     /db_xref="GI:14025723"
                     /translation="MHVVADRSETLTAIRTDLGAIFVSLELSRSTWLITSLSPGGGEK
                     LSKHSVPSGDLSALFRRFDSVREKARARMGRDFPVVTIQEAGLDGFWIHRVLSKAGIE
                     SYVVDPASIAVSRRSRRAKTDKIDGETLVRSLLAFKRGEPRICAMVKEPTPEQEDRRR
                     ICRERKVLTTERVCHVNRIKGLLFAQGITDYEPLNRDRRDRLEELVTGDGRSLPPHLK
                     AQISRELDRLELLLAQVKEIEIERDKLLHAEEKVIHSPAKLLIDIKGIGPEFAAVLCH
                     EGLFRHFDNRRQVAAYAGLAPTPWQSGQVDREQGVSKAGNPRLRSTMVQLAWLWIRHQ
                     PRSNLTLWFQSRLRRNGGRLKKTTIVALARKLLIALWKYATSGVIIEGALMKAA"
     gene            3811. .4977
                     /gene="mlr5965"
     CDS             3811. .4977
                     /gene="mlr5965"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52323.1"
                     /db_xref="GI:14025724"
                     /translation="MGHSLLFAVDVRPPHDVGVDGPRTASMCQNEVVADLKQRRPSVE
                     QIIRIGIDTSKHVFQLHGVNAAEEPVLRKKLRRKEVVALFEKLSPTVVAIEACGGSHH
                     WARLLQSFGHEVKLIPPQYVKPYVKRGKNDTADAEALCEAMSRPTMRFVPVKTAKQQA
                     ALMLVGLRDRLIRNRTQLANAIRGYAAEYGLIAARGMCKIEPLLERIAADKMLPDLAR
                     ELFALHAKEYAQLQTQLKDVDAKLMAWHRAEACSQRLAQIPGVGPIGASLLVMKTPAP
```

```
                ETFRSARHFAAWLGLTPKDHSTAGRVRLGVITRAGDEALRSVLVAGATAVIRHVRRGR
                GAVSPWLVDLLKRKPPKLAAVALANKIARIAWKLMVSGEIYGAKPMPPASARAA"
gene            complement(5434. .5571)
                /gene="msl5966"
CDS             complement(5434. .5571)
                /gene="msl5966"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB52324.1"
                /db_xref="GI:14025725"
                /translation="MNIAGRRGMARARAAVARKLAVILHRMWADATGFRFGKEPVHLA
                A"
gene            complement(5568. .6362)
                /gene="mll5967"
CDS             complement(5568. .6362)
                /gene="mll5967"
                /codon_start=1
                /transl_table=11
                /product="transposase"
                /protein_id="BAB52325.1"
                /db_xref="GI:14025726"
                /translation="MKRALRGTGLPLERIGLEACSLRAWLHDVLHAEGLPATCIEMRQ
                ANAAMKTMPNKTDRNDARALAQIMRTPGIAWCMSRAAGSAGCGVRCLVARRTVLNEMR
                SIENVVQAILREAGVKLGLPSRTTFANRVRELTGADAEVMAMVEPLLAILGVMLREFA
                RLTKQVLDIVRNEKVCRRLMSAPGVGPITALAFRATIDRPERFGSSQAVGAHLGLTPA
                RYSPAKPTSRERSADAATNLYAPRFMRRPTHCWCDRRSGRASEHGA"
gene            6679. .7272
                /gene="mlr5968"
CDS             6679. .7272
                /gene="mlr5968"
                /codon_start=1
                /transl_table=11
                /product="transposase"
                /protein_id="BAB52326.1"
                /db_xref="GI:14025727"
                /translation="MGWTEATRRQYCRAGQRDANALTDREWAAIEPSMPVAKATGRPR
                TTELRAVMDALLYIAWTGCQWRALPDRFPPVSTVKRYFYAWRDNGLWKSINFHLVAAA
                RLALGREASPSAGIIDSQSAKATDVAGLRGYDAGKKIKGRNAISSPTPRAPGRAAGAY
                GRHPGEGWGRRRHRLHSTALPVVAAICSPTVVMAEKN"
gene            6948. .7511
                /gene="msr8681"
CDS             6948. .7511
                /gene="msr8681"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB52327.1"
                /db_xref="GI:14025728"
                /translation="MEEHQLPSGGRGTSGAWARGKPERRHHRQPIGKGHGRCRFAWLR
                RRQEDQRPQRHIITDTEGTWSGCRCIRPTSRRGMGQAASSPPFDSFARGCRHLFADGG
                YGGEKLTHALADPGTWTIEIIKRSDTAKGFEVLPRRWVVERTFAWLGRCRRLAKDFEA
                TIASAEAWIFIASIRLVLRRMAAPRRA"
gene            7627. .7719
                /gene="msr5969"
CDS             7627. .7719
                /gene="msr5969"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB52328.1"
                /db_xref="GI:14025729"
                /translation="MGASKGERTPERLGYLSGYYGRTLVTRRQA"
gene            complement(7860. .9029)
                /gene="mll5970"
CDS             complement(7860. .9029)
                /gene="mll5970"
                /codon_start=1
                /transl_table=11
                /product="putative peptidase"
                /protein_id="BAB52329.1"
                /db_xref="GI:14025730"
                /translation="MPIPKGLQAFPRAEYLRRLSAVKAEMGRRETNTLIVTTSADITY
                LSGYTCKSGYVPQGLIISLAEEEPTFVARRCDAPAAMHQMFIDSGRVIGYPEALIANP
                RADGFDAMIDFLRQNGLAKGRIGLETKSLRSETVEKFKARMSAVAIVDFSNTIAWIRL
                AKSDLEIAVMREAAAITDAGMLRAKEVIRPGVREADAAAEIISTLIRGADGKPGTDLA
                GFYLCASPRSATAHVHWSEDVFRDGSQINLEFGGVRHGYVSGLMRTFSIGEPSDRLRR
                VHEAEVAGLESALAAVKLGVTCGDIATAFNKTLRRHGFDKETRCGYAIGIDWLEPTAS
                LKENDETVLQTNMTFHLMLGNWIDEDFGYVISETFRVTERGAETFSCLPREIFEI"
gene            complement(9985. .10185)
                /gene="msl8683"
CDS             complement(9985. .10185)
                /gene="msl8683"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB52330.1"
                /db_xref="GI:14025731"
                /translation="MMIDLTSADDTRLKQVELSSAIHLAFHELELSDLALGLTVGPWQ
                GDRRADGQTVFENAVGERGDQT"
gene            10610. .11893
                /gene="mlr5971"
CDS             10610. .11893
                /gene="mlr5971"
                /codon_start=1
                /transl_table=11
                /product="L-lysine 2,3-aminomutase"
                /protein_id="BAB52331.1"
                /db_xref="GI:14025732"
                /translation="MSRIAHETMRHEFADAAWQDHHAVTRVDQLREWMGADADGNLIR
                DVERGLARAPMALRITPYLLNLIDWSNFLEDPIRKQFIPVGSELLPSHPLLKMDSLHE
                RKSSPVDGLVHRYKDKVLLLATDRCPVYCRFCTRSYSVGLDTQSVLKKKVSPFQSRWD
                TILAYLRVTPVIADVVVSGGDCFRLKPSRLLAIGMGLLSIPSIRRIRFATKGLAVLPM
                KITSDHKWTDALVNISDAGRDQGVEISFHTHFNHPREITDYTIAAAELLFKRGIRMRN
                QSVLMAGVNDDPEVMKQLVKKLSDLHIQPYYVYTCDLVDGIEHMRCSVRLACQIEKAV
                RGITAGYNTPLFVVDTPGGGGKRDVHSFEHYNTQTGLAVYRALSVRPDQLFTYPDPLQ
                SLSPDAQAAWVDPWRAKGMLREAVRAARVKLMGLE"
gene            12167. .12463
                /gene="msr5972"
CDS             12167. .12463
                /gene="msr5972"
                /codon_start=1
                /transl_table=11
                /product="transposase"
                /protein_id="BAB52332.1"
                /db_xref="GI:14025733"
                /translation="MGMQRRKFSREFKLEAVRLIKDRGVAVAQAARDLDVHENVLRKW
                VRDLSTDPQHAFPGHGQMKPEQLEIDRLRKEVAKLKAERDILKKAAAYFAREVT"
gene            12460. .13323

alignment_scores:
            Quality:  100.00              Length:    232
              Ratio:  0.855                 Gaps:      9
 Percent Similarity:  50.431   Percent Identity: 19.828

alignment_block:
 US-09-674-779-2 x AP003008/rev   ..

 Align seg 1/1  to reverse of: AP003008  from: 1  to: 346820

       2 LysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuVa 18
         :::::::::|||   |||:::|||:::   ::::::::::::
   21089 GAATCTATCAATGCATATTTGATCGCCCGTTTCGTGACCCAA..... 21048

      18 lAlaCysSerAlaProIleProThrAsnProGlnVal...SerProIleL 34
               :::   |||   |||         ::::::   :::   ::::
   21047 .......GGCGAGCCTTAGCCGCAGCCAGCCCGGATGGCTGCAGACGTCA 21005

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         ::      :::|||||:::   :::   :::   |||   :::
   21004 ATTGCGCCAAAGTTTTATTCATCTCGAGGCGGGAGGGA...CAGATTTTG 20958

      51 HisGluHisAspGluSerValSerHisValGlyLeu......GlnAlaHi 65
         :::   :::|||   :::|||:::   |||   |||       |||   ||
   20957 AACACTCGTGACGCAACAGTGGCTGATGTGGACCTTATCGCGCAGCTTCA 20908
```

```
    65 sPheGluThrTrp.....................LeuGlnMetHisHisA 75
       |:::|||:::|||                     |||:::      :::
 20907 TATTGAAAGTTGGCGCAACGCCTACGCAGGGATACTCTCGGCATCCTATC. 20858

    75 laThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeu 91
         :::         |||:::     |||:::     ::::::|||:::
 20857 TAGCGGGTCCGATAGTGCTCGACCGTCAGACTGTCTGGCGAGAGCGTTTC 20808

    92 GlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTr 108
       ::::::      |||   :::   ::::::::::::::   ::::::
 20807 AATTCACCGGCTCCAAATCTC...AAAGTCATAGTTGCCGAAGAAGAAGG 20761

   108 pGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpG 125
        ||||||   |||                  |||      ||||
 20760 CCAAGCGCTGGGCTTCGTCTGCCTCTTAGGCAATGCGGAGGACCGTTGGG 20711

   125 lyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGly 141
       ||   :::|||   :::||||||:::           |||:::||||||
 20710 GTACCCTCGTCGACAATCTCCATGTTTTGCCAACGGCGAAGGGGCGGGGC 20661

   142 IleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLe 158
       :::                                 :::   :::||
 20660 GTC................................GGAAGGCATCT 20647

   158 uAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerA 175
       |   :::   |||:::|||   :::   :::::::::      :::   :
 20646 TATTCGCGTTGCTGCGGGCTGGTCA...GCCGAAAATTACCCGGGCGTGG 20600

   175 laIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr 191
       :::::   :::||||||   :::::::::      ::::::|||:::|||
 20599 GGCTGCATCTATGGGTCTACGAGGTTAATGCCCCTGCGCGCGCTTTCTAC 20550

   192 Glu..................................................  192
       |||
 20549 GAGAGAATGGGTGGTCAGGTAGTCGCGCGCCTACCGCAGTCCAACCCTGA 20500

   193 .......LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGly 205
             ::::::   :::|||||||   ||||||   :::   |||
 20499 CGGAAGAATCCACGTCGAACTGTGCTACTATTGGCCTGATTCTGGC 20454
```

```
seq_name: gb_htg:AC105039

seq_documentation_block:
LOCUS       AC105039              145667 bp    DNA     linear   HTG 10-FEB-2002
DEFINITION  Homo sapiens chromosome 18 clone CTD-2357J19 map 18, WORKING DRAFT
            SEQUENCE, 4 ordered pieces.
ACCESSION   AC105039
VERSION     AC105039.3  GI:18643649
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145667)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone CTD-2357J19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 145667)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 10, 2002 this sequence version replaced gi:18598788.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23024
                Center clone name: 2357_J_19
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 145044 bases at least Q40
                Consensus quality: 145204 bases at least Q30
                Consensus quality: 145258 bases at least Q20
                Insert size: 145000; agarose-fp
                Insert size: 145367; sum-of-contigs
                Quality coverage: 12.1 in Q20 bases; agarose-fp
                Quality coverage: 12.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1       75: contig of 75 bp in length
            *      76 175: gap of     100 bp
            *     176    13700: contig of 13525 bp in length
            *   13701 13800: gap of     100 bp
            *   13801    46229: contig of 32429 bp in length
            *   46230 46329: gap of     100 bp
            *   46330   145667: contig of 99338 bp in length.
FEATURES             Location/Qualifiers
     source          1. .145667
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="CTD-2357J19"
                     /clone_lib="CITD Human BAC"
     misc_feature    1. .75
                     /note="assembly_fragment"
     misc_feature    176. .13700
                     /note="assembly_fragment"
     misc_feature    13801. .46229
                     /note="assembly_fragment"
     misc_feature    46330. .145667
                     /note="assembly_fragment"
BASE COUNT    45215 a  27244 c  27442 g  45464 t    302 others
ORIGIN


alignment_scores:
           Quality:   99.50            Length:    141
             Ratio:  1.345              Gaps:      5
```

```
 Percent Similarity:  52.482   Percent Identity: 24.823

alignment_block:
 US-09-674-779-2 x AC105039/rev   ..

 Align seg 1/1  to reverse of: AC105039  from: 1  to: 145667

       9 IleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIlePr 25
         ::::|||:::|||   :::   :::|||::::::|||      :::
   89241 CTAACTTCTTTGTATAACCGCCTTCTCCTCTCTTGCTTCTTTATACTTGA 89192

      25 o.................ThrAsnProGlnValSerProIleLysThrP 36
                           :::|||   :::      |||:::   ||||
   89191 CTATTAAACTACAACTATAGTTAATACCAAATCTCTACCTATGGCTACTC 89142

      36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGlu 52
         ||:::   |||         ||||||   |||:::   |||||||||
   89141 CCACACAATTG.........GATAAGGCTGGAGAAACACACACACACACA 89101

      53 HisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTr 69
         |||   ::::::   :::|||:::      :::   |||   |||
   89100 CACACACACACACACACACACACACACACACACACAAACACACAGAATGTAT 89051

      69 pLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaT 86
          |||::::::|||   |||      ::::::::::::   |||
   89050 CCTTCATCTTCACCTAGCT......GATGACCTTGCTTCCTATTTTCTTA 89007

      86 yrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu 102
           :::|||||||:::|||   :::
   89006 CTTTCCAAAGCAAACTGAAACAC.......................... 88984

     103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
            :::   :::|||::::::   |||   :::      ::::::|||||
   88983 ...TCAGAAGAGAGCTTCCACAGATGCCCACAACTATATCACTCCTTACC 88937

     119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136
         |                 :::   |||      |||   ::::::   :
   88936 T.................TCCTGTGTCATGTTCCTCTCTATGTTCCTCT 88905

     136 spLeuLysSerArgGlyIleLeu 143
         :::::   |||   :::|||||||
   88904 CTATGTTCAGTCCCTCTATCCTT 88882

seq_name: gb_htg:AC016944

seq_documentation_block:
LOCUS       AC016944              158886 bp    DNA     linear   HTG 27-AUG-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-70P20, WORKING DRAFT SEQUENCE,
            11 unordered pieces.
ACCESSION   AC016944
VERSION     AC016944.14  GI:9929553
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158886)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158886)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 26, 2000 this sequence version replaced gi:9719630.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: HMUF
                Center clone name: RP11-70P20
            ---------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 144562 bases at least Q40
                Consensus quality: 152026 bases at least Q30
                Consensus quality: 154719 bases at least Q20
                Estimated insert size: 155752; sum-of-contigs estimation
                Estimated insert size: 209830; agarose-fp estimation
                Quality coverage: 3.2x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.3x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    43752: contig of 43752 bp in length
            *    43753    43852: gap of unknown length
            *    43853    71719: contig of 27867 bp in length
            *    71720    71819: gap of unknown length
            *    71820    93367: contig of 21548 bp in length
            *    93368    93467: gap of unknown length
            *    93468   108864: contig of 15397 bp in length
            *   108865   108964: gap of unknown length
            *   108965   124124: contig of 15160 bp in length
            *   124125   124224: gap of unknown length
            *   124225   135573: contig of 11349 bp in length
            *   135574   135673: gap of unknown length
            *   135674   141767: contig of 6094 bp in length
            *   141768   141867: gap of unknown length
            *   141868   149244: contig of 7377 bp in length
            *   149245   149344: gap of unknown length
            *   149345   155749: contig of 6405 bp in length
            *   155750   155849: gap of unknown length
            *   155850   157240: contig of 1391 bp in length
            *   157241   157340: gap of unknown length
            *   157341   158886: contig of 1546 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158886
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-70P20"
BASE COUNT    45757 a  33003 c  33199 g  45917 t   1010 others
ORIGIN
```

```
alignment_scores:
            Quality:   99.50              Length:    270
              Ratio:  0.783                Gaps:     14
 Percent Similarity:  47.037   Percent Identity: 18.889

alignment_block:
 US-09-674-779-2 x AC016944/rev   ..

 Align seg 1/1  to reverse of: AC016944  from: 1  to: 158886

      28 ProGlnValSerProIleLysThrProSerValLeuIleThrLysAspLy 44
         |||::::::::::|||   :::   |||||||||   :::
  145016 CCCTCCATCCAGCCTCCCCAGCCCCCATCTGTCTGTCTCTGT........ 144975

      44 sIleGlyAspHisHisThrHisGluHisAspGluSerValSerHisVal. 60
          :::    |||||||||   |||   ::::::   :::|||:::
  144974 .CTTCCTTACACACACACACACGCACACACACACACACACACACATGCTT 144926

      61 ....................GlyLeuGlnAlaHisPheGluThrTrpLeu 70
                               :::|||:::   :::      |||
  144925 AGGAGAAACAGGAACCAGAGGAGATTCAGACATTTATTTTGCTTTGG... 144879

      71 GlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLe 87
         |||:::::::|||      |||:::      :::   :::          ||
  144878 CAAATCTACCACTGACCTAAGAAGCCAGGACTTTTGTTTGCT......TT 144835

      87 uGlnSerArgLeuGly............AsnTyrLeuProProMetSer. 99
         |:::|||:::::|||            :::||||||      :::|||
  144834 AGAATCTCAGATGGGAAAAGGCACCCTAAGTTATTTATTTGTTTTTTCAT 144785

     100 ..............GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys 111
                      :::::::|||   ::::::::::::|||||||||||
  144784 GAATTCTTTTTTATTCTGTACTTAGAGTCAGTCAGGGATGGCAAGCATGT 144735

     112 Gly.....................................His..... 113
         |||                                    |||
  144734 GGTGTCGGTCTGTAGGGCAGCTGTGTACATCTGGGAGACATTCATAGTGT 144685

     114 .GluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIle....... 127
          :::      ::::::   |||:::|||      :::::::::
  144684 GAGCACTGACTCTGTCATTCCCCAGCATAGAAGGAGCAATGTAGCTGGTG 144635

     128 ........................ValProThrLeuHisLeuTyrGln 135
                                  :::|||   :::        :::
  144634 CTTCTGTGCCCCTATGAGCAGCCCAAATTCCAAAAATTCCATGGGAAGAA 144585

     136 Asp......LeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleAr 150
                 |||   ::::::   ::::::|||:::::::::::::   ||
  144584 GGGACACGGCTGGCTGAGCACACTGTGATTCCCACCACAGCTGAGTGCAG 144535

     150 gSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaM 167
         |:::                                        ::::
  144534 GGAT.....................................CCTGGGG 144524

     167 etSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeu 183
         :::::::::::::::::|||   ::::::      :::      |||
  144523 TGAAAAATGAGATTGTAAACATCACCCTGGCAGCATTTTCATCAGATGAA 144474

     184 GluIleLysSer............GlnAlaLeuTyrGluLeuGlnAsnAr 196
         |||::::::::            ::::::   :::|||   |||:::::
  144473 GAGATGGAAGGCCTGCACGATCTCAAGGGAGCCTGGGAAAATCAACACAA 144424

     196 gLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuG 213
         :   |||   |||:::|||:::   |||:::|||:::|||
  144423 TGCCTGCGTATATTTTCTAGAT...GGGCAAAATAAAAAC.......... 144387

     213 lyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLys 229
                                                     |||
  144386 ..........................................AAA 144384

     230 TrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuPr 246
               :::|||:::|||   ::::::      :::|||:::
  144383 CCAAAAAAGAGGTTTAAGGAAGACGATGATGCAGGAAAGCACATCAGCAC 144334

      246 oLysAsnLys 249
              ||||||
  144333 AATAAACAAA 144324

seq_name: gb_pr:AL139038

seq_documentation_block:
LOCUS       AL139038              140756 bp    DNA     linear   PRI 15-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-456B18 on chromosome 13,
            complete sequence.
ACCESSION   AL139038
VERSION     AL139038.18  GI:14800148
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140756)
  AUTHORS   Dunn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 17, 2001 this sequence version replaced gi:13751261.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-456B18 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-456B18 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-456B18 is at 140756 in this
            sequence. The true left end of clone RP11-78L16 is at 68441 in this
            sequence. The true right end of clone RP11-108H9 is at 100 in this
            sequence.
FEATURES              Location/Qualifiers
     source           1. .140756
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="13"
                      /clone="RP11-456B18"
                      /clone_lib="RPCI-11.2"
     repeat_region    1830. .2326
                      /note="LOR1a repeat: matches 1. .497 of consensus"
     repeat_region    3932. .3965
                      /note="17 copies 2 mer gt 94% conserved"
     repeat_region    4632. .5067
                      /note="LOR1b repeat: matches 1. .461 of consensus"
     repeat_region    5733. .5893
                      /note="L1ME3 repeat: matches 6002. .6146 of consensus"
```

```
repeat_region   6000. .6083
                /note="MER34 repeat: matches 454. .539 of consensus"
repeat_region   6108. .6171
                /note="HERV23 repeat: matches 993. .1056 of consensus"
repeat_region   7143. .7840
                /note="LTR8 repeat: matches 1. .691 of consensus"
repeat_region   8059. .8292
                /note="LTR16A repeat: matches 203. .442 of consensus"
repeat_region   8491. .8686
                /note="LTR29 repeat: matches 6. .203 of consensus"
repeat_region   8725. .8926
                /note="MER4D repeat: matches 483. .677 of consensus"
repeat_region   8992. .9049
                /note="MER4D repeat: matches 410. .465 of consensus"
repeat_region   9062. .9281
                /note="110 copies 2 mer tt 59% conserved"
repeat_region   9398. .9530
                /note="MER41C repeat: matches 1. .131 of consensus"
repeat_region   9623. .9732
                /note="LTR1 repeat: matches 675. .785 of consensus"
repeat_region   9970. .10333
                /note="MLT2D repeat: matches 1. .486 of consensus"
repeat_region   10400. .10545
                /note="LTR29 repeat: matches 2. .151 of consensus"
repeat_region   10549. .10935
                /note="MER67D repeat: matches 12. .391 of consensus"
repeat_region   11249. .11332
                /note="MIR repeat: matches 69. .144 of consensus"
repeat_region   11923. .12159
                /note="MER4D repeat: matches 1. .234 of consensus"
repeat_region   12164. .12767
                /note="MER4D repeat: matches 396. .973 of consensus"
repeat_region   12852. .13221
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   13222. .14798
                /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                consensus"
repeat_region   14799. .15173
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   16144. .16346
                /note="MER54B repeat: matches 585. .789 of consensus"
repeat_region   16339. .16503
                /note="MER54B repeat: matches 112. .278 of consensus"
repeat_region   16504. .16571
                /note="MER66A repeat: matches 1. .69 of consensus"
repeat_region   16570. .16629
                /note="MER41B repeat: matches 283. .341 of consensus"
repeat_region   16599. .16902
                /note="MER66A repeat: matches 163. .478 of consensus"
repeat_region   16903. .17014
                /note="MER54B repeat: matches 3. .115 of consensus"
repeat_region   17195. .17491
                /note="AluY repeat: matches 1. .297 of consensus"
repeat_region   18120. .19291
                /note="L1P repeat: matches 1726. .2897 of consensus"
repeat_region   19460. .19584
                /note="25 copies 5 mer atttt 62% conserved"
repeat_region   19686. .19745
                /note="12 copies 5 mer tttaa 71% conserved"
repeat_region   20453. .20914
                /note="L1MD2 repeat: matches 5745. .6199 of consensus"
repeat_region   20913. .21222
                /note="L1MD2 repeat: matches 5277. .5594 of consensus"
repeat_region   21308. .21570
                /note="LTR9 repeat: matches 357. .625 of consensus"
repeat_region   21645. .21984
                /note="LTR19B repeat: matches 55. .393 of consensus"
repeat_region   22081. .22583
                /note="L1MD2 repeat: matches 4726. .5242 of consensus"
repeat_region   23498. .23784
                /note="AluJb repeat: matches 1. .305 of consensus"
repeat_region   23873. .24169
                /note="AluSg repeat: matches 1. .295 of consensus"
repeat_region   25044. .25238
                /note="MIR repeat: matches 31. .238 of consensus"
repeat_region   25404. .25696
                /note="AluSx repeat: matches 20. .311 of consensus"
repeat_region   25832. .26218
                /note="MSTC repeat: matches 1. .405 of consensus"
repeat_region   26221. .27865
                /note="MSTC-internal repeat: matches 2. .1651 of consensus"
repeat_region   27907. .28659
                /note="L1PB2 repeat: matches 5400. .6155 of consensus"
repeat_region   28710. .29043
                /note="MSTD repeat: matches 30. .394 of consensus"
repeat_region   29634. .29697
                /note="32 copies 2 mer ta 79% conserved"
repeat_region   29729. .29752
                /note="12 copies 2 mer tg 95% conserved"
repeat_region   30688. .30874
                /note="MER61A repeat: matches 160. .354 of consensus"
repeat_region   31648. .31785
                /note="L1MC4 repeat: matches 6409. .6540 of consensus"
repeat_region   31813. .32243
                /note="L1MC4 repeat: matches 6629. .7089 of consensus"
repeat_region   32449. .32821
                /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region   32824. .34464
                /note="THE1C-internal repeat: matches 3. .1651 of
                consensus"
repeat_region   34465. .34825
                /note="THE1C repeat: matches 1. .360 of consensus"
repeat_region   34804. .35004
                /note="MER67C repeat: matches 202. .403 of consensus"
repeat_region   39792. .39868
                /note="L1PA12 repeat: matches 6076. .6152 of consensus"
repeat_region   40953. .41258
                /note="AluY repeat: matches 1. .308 of consensus"
repeat_region   41279. .41308
                /note="15 copies 2 mer aa 86% conserved"
repeat_region   42744. .42933
                /note="L1MA8 repeat: matches 6064. .6256 of consensus"
repeat_region   42973. .43275
                /note="AluSx repeat: matches 1. .303 of consensus"
misc_feature    complement(43560. .43991)
                /note="match: STS: Em:HSPH07E5"
repeat_region   45196. .45584
                /note="MLT2B repeat: matches 2. .399 of consensus"
repeat_region   45594. .45767
                /note="87 copies 2 mer ta 77% conserved"
repeat_region   45800. .45851
                /note="MLT2B repeat: matches 394. .448 of consensus"
repeat_region   48118. .48487
                /note="THE1C repeat: matches 1. .370 of consensus"
repeat_region   48708. .48806
                /note="AluJo/FRAM repeat: matches 197. .297 of consensus"
repeat_region   49853. .50271
                /note="L1MEc repeat: matches 2274. .2369 of consensus"
repeat_region   50637. .50927
                /note="L1ME repeat: matches 974. .1253 of consensus"
repeat_region   52250. .52349
                /note="MIR repeat: matches 131. .228 of consensus"
repeat_region   52696. .52996
                /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region   54927. .55010

alignment_scores:
            Quality:   99.00              Length:    183
              Ratio:  1.138                Gaps:      8
 Percent Similarity:  47.541   Percent Identity: 25.683

alignment_block:
 US-09-674-779-2 x AL139038/rev   ..

 Align seg 1/1  to reverse of: AL139038  from: 1  to: 140756
```

```
    34 LysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisTh 50
33228 AAAAAGCCACCAGTCTCTTTGCTAAAGCATAGCAAGACTGACTTTTACTC 33179

    50 rHisGluHisAspGluSerValSerHisValGlyLeuGlnAla..... 64
33178 C......AGTTCCCAACAAGTTTCTCATCTCCATCTGAGACCACCTCAGT 33135

    65 .....HisPheGluThrTrpLeuGlnMetHisHis 74
33134 CTGGAATTCATTTTCCATATCATTATCAGCCTTTTGGTCAAAGCCATTCA 33085

    75 AlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg.. 90
33084 ACAAATCTCTAGGAAATTCCAAACTTTACCACATACCTGTCTTCTGAGCC 33035

    91 ..........LeuGlyAsn............TyrLeuProProMetSerG 100
33034 CTCCAAGTCTCTAGGAAATTCCATACTTTTTCACATTTTCCCGTCTTCTG 32985

   100 lnLeuLeuThrThrAlaArgSerTrpGln..............AlaCys 111
32984 AGCTTTCCAAGTCTCTAAGAAGTTCCAAACTTTCCTACATTTTCCTGTGT 32935

   112 GlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleVa 128
32934 TCTTCTGAGCCCTCCAAACTGTTCCAACCTCTGCCTGTTACTCAATTCCA 32885

   128 lProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProA 145
32884 AAGCTGCTTCCACATTTTTGAGTATCTTTACAGCAGAGCCATACTA.... 32839

   145 laAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCys 161
32838 .................TCCTGGTACCAATTTACTGTATTA......GTC 32812

   162 AlaGlyGlyAlaAlaMetSerLysHis..................... 170
32811 TGTTCTCATGCCGCTATGAATAAATACCTGAGACTGGGTAATTAAAAAAG 32762

   171 ................LeuThrAsnSerAlaIleAspIleTrpValPro 181
32761 AAACAAAAAGGTTTAATTGACTCACAGTTTTGCAGGGCTGTGGGGGCCT 32713
```

seq_name: gb_htg:AC073525

seq_documentation_block:

LOCUS AC073525 196048 bp DNA linear HTG 04-FEB-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-81K13, WORKING DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION AC073525
VERSION AC073525.20 GI:18481944
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 196048)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196048)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Feb 3, 2002 this sequence version replaced gi:15799494.

```
---------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
    Center project name: HBGM
    Center clone name: RP11-81K13
---------- Summary Statistics
    Sequencing vector: Plasmid; M77789
    Sequencing vector: M13; L08821
    Chemistry: Dye-primer Bodipy: 32% of reads
    Chemistry: Dye-terminator Big Dye: 68% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 226075 bases at least Q40
    Consensus quality: 237688 bases at least Q30
    Consensus quality: 247848 bases at least Q20
    Estimated insert size: 252844; sum-of-contigs estimation
    Quality coverage: 0x in Q20 bases; agarose-fp estimation
    Quality coverage: 5.3x in Q20 bases; sum-of-contigs estimation
----------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1   142460: contig of 142460 bp in length
*  142461   142560: gap of unknown length
*  142561   145652: contig of 3092 bp in length
*  145653   145752: gap of unknown length
*  145753   150236: contig of 4484 bp in length
*  150237   150336: gap of unknown length
*  150337   153577: contig of 3241 bp in length
```

```
            *   153578   153677: gap of unknown length
            *   153678   155916: contig of 2239 bp in length
            *   155917   156016: gap of unknown length
            *   156017   158716: contig of 2700 bp in length
            *   158717   158816: gap of unknown length
            *   158817   162262: contig of 3446 bp in length
            *   162263   162362: gap of unknown length
            *   162363   164458: contig of 2096 bp in length
            *   164459   164558: gap of unknown length
            *   164559   168040: contig of 3482 bp in length
            *   168041   168140: gap of unknown length
            *   168141   170418: contig of 2278 bp in length
            *   170419   170518: gap of unknown length
            *   170519   172921: contig of 2403 bp in length
            *   172922   173021: gap of unknown length
            *   173022   175106: contig of 2085 bp in length
            *   175107   175206: gap of unknown length
            *   175207   177873: contig of 2667 bp in length
            *   177874   177973: gap of unknown length
            *   177974   180687: contig of 2714 bp in length
            *   180688   180787: gap of unknown length
            *   180788   183385: contig of 2598 bp in length
            *   183386   183485: gap of unknown length
            *   183486   186345: contig of 2860 bp in length
            *   186346   186445: gap of unknown length
            *   186446   188906: contig of 2461 bp in length
            *   188907   189006: gap of unknown length
            *   189007   191567: contig of 2561 bp in length
            *   191568   191667: gap of unknown length
            *   191668   193864: contig of 2197 bp in length
            *   193865   193964: gap of unknown length
            *   193965   196048: contig of 2084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196048
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-81K13"
BASE COUNT    62213 a  35817 c  36336 g  59703 t   1979 others
ORIGIN


alignment_scores:
           Quality:   99.00              Length:    183
             Ratio:  1.138                Gaps:      8
 Percent Similarity:  47.541   Percent Identity: 25.683

alignment_block:
 US-09-674-779-2 x AC073525/rev   ..

 Align seg 1/1  to reverse of: AC073525  from: 1  to: 196048

      34 LysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisTh 50
         |||   |||   |||   :::   |||   :::      |||   :::::
  143398 AAAAAGCCACCAGTCTCTTTGCTAAAGCATAGCAAGACTGACTTTTACTC 143349

      50 rHisGluHisAspGluSerValSerHisValGlyLeuGlnAla....... 64
         :        :::::::::|||||||||:::   |||:::
  143348 C......AGTTCCCAACAAGTTTCTCATCTCCATCTGAGACCACCTCAGT 143305

      65 ....................HisPheGluThrTrpLeuGlnMetHisHis 74
                             |||::::::   :::   |||   |||
  143304 CTGGAATTCATTTTCCATATCATTATCAGCCTTTTGGTCAAAGCCATTCA 143255

      75 AlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg.. 90
         ::::::      |||:::   ::::::   :::||||||:::|||:::
  143254 ACAAATCTCTAGGAAATTCCAAACTTTACCACATACCTGTCTTCTGAGCC 143205

      91 ..........LeuGlyAsn............TyrLeuProProMetSerG 100
                 |||||||||            ::::::   |||   |||:
  143204 CTCCAAGTCTCTAGGAAATTCCATACTTTTTCACATTTTCCCGTCTTCTG 143155

     100 lnLeuLeuThrThrAlaArgSerTrpGln..............AlaCys 111
         ::|||      :::   ||||||   :::                  |||
  143154 AGCTTTCCAAGTCTCTAAGAAGTTCCAAACTTTCCTACATTTTCCTGTGT 143105

     112 GlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleVa 128
         :::   ||||||   :::|||                  |||:::
  143104 TCTTCTGAGCCCTCCAAACTGTTCCAACCTCTGCCTGTTACTCAATTCCA 143055

     128 lProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProA 145
             :::|||:::::::::   |||   ||||||:::||||||
  143054 AAGCTGCTTCCACATTTTTGAGTATCTTTACAGCAGAGCCATACTA.... 143009

     145 laAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCys 161
                         |||   |||:::         |||
  143008 ................TCCTGGTACCAATTTACTGTATTA......GTC 142982

     162 AlaGlyGlyAlaAlaMetSerLysHis..................... 170
         ::::::   |||||||||:::|||:::
  142981 TGTTCTCATGCCGCTATGAATAAATACCTGAGACTGGGTAATTAAAAAAG 142932

     171 ..............LeuThrAsnSerAlaIleAspIleTrpValPro 181
                         ||||||:::|||         :::|||   |||
  142931 AAACAAAAAGGTTTAATTGACTCACAGTTTTGCAGGGCTGTGGGGGCCT 142883

seq_name: gb_htg:AC103346

seq_documentation_block:
LOCUS       AC103346              30062 bp    DNA     linear   HTG 24-NOV-2001
DEFINITION  Rattus norvegicus clone CH230-2K7, *** SEQUENCING IN PROGRESS ***,
            9 unordered pieces.
ACCESSION   AC103346
VERSION     AC103346.1  GI:17063114
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 30062)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
```

```
          Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
          Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
          Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 30062)
  AUTHORS Worley,K.C.
  TITLE   Direct Submission
  JOURNAL Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   ---------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          ---------- Project Information
              Center project name: TUVO
              Center clone name: CH230-2K7
          ---------- Summary Statistics
              Sequencing vector: Plasmid; M77789
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 68432 bases at least Q40
              Consensus quality: 72999 bases at least Q30
              Consensus quality: 77137 bases at least Q20
              Estimated insert size: 61291; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
          ----------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     6200: contig of 6200 bp in length
          *     6201     6300: gap of unknown length
          *     6301     9304: contig of 3004 bp in length
          *     9305     9404: gap of unknown length
          *     9405    13486: contig of 4082 bp in length
          *    13487    13586: gap of unknown length
          *    13587    17093: contig of 3507 bp in length
          *    17094    17193: gap of unknown length
          *    17194    19456: contig of 2263 bp in length
          *    19457    19556: gap of unknown length
          *    19557    22347: contig of 2791 bp in length
          *    22348    22447: gap of unknown length
          *    22448    24839: contig of 2392 bp in length
          *    24840    24939: gap of unknown length
          *    24940    27216: contig of 2277 bp in length
          *    27217    27316: gap of unknown length
          *    27317    30062: contig of 2746 bp in length.
FEATURES             Location/Qualifiers
     source          1. .30062
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-2K7"
BASE COUNT     7816 a   6888 c   6714 g   7802 t    842 others
ORIGIN


alignment_scores:
            Quality:   98.50            Length:    165
              Ratio:  1.145              Gaps:      8
 Percent Similarity:  52.121   Percent Identity: 26.667

alignment_block:
```

```
US-09-674-779-2 x AC103346   ..

Align seg 1/1  to: AC103346  from: 1  to: 30062

     21 SerAlaProIleProThrAsnProGlnValSerProIleLysThrProSe 37
        :::|||      |||::::::   ::::::|||            |||||
  14376 GCTGCTGTGTGCCCAGCCAGTGACAATACCAGCCAAGCTGTTCAGCCTTC 14425

     37 rValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisA 54
        |:::  :::::::::              ||||||||||||   :::
  14426 CCTCTAGGTAAGCAGTGGTTTTTACACACCACACCACACGCACGCACGCA 14475

     54 spGluSerValSerHisValGlyLeuGlnAlaHisPheGlu..... 67
        ::::::   :::|||:::      :::|||:::   :::
  14476 CGCACGCACGCACGCACGCACGCACGCACGCGCGCGCCAGCTCCAGACAG 14525

     68 .....ThrTrpLeuGlnMetHis...HisAlaTh 76
                    |||||||||:::::::::   |||:::
  14526 AGTAAATAATGATGGAGGGCCACATGGTTGAGAGTTGAAGTGCATTGTTT 14575

     76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
        :::         :::|||:::           |||||||||:::|||:
  14576 CAGAGCTGGAGAGTTGAGATGG.....CAGAGCAGAATAGGTT 14613

     93 snTyrLeuProProMetSerGlnLeuLeuThrThr.AlaArgSerTrpGl 109
        ::   :::      |||   |||     |||    :::|||:::|||
  14614 CAGGAATAGATTGTATGATGCAG......ACTTTGAAGCAGACAATGGCT 14657

    109 nAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyG 126
                       :::|||||||||   |||||||:::
  14658 TCTT.............GTGTCCCTCCCACCTGGCCATTTGTTT.... 14688

    126 lnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIle 142
                     |||   |||||||||:::||||||
  14689 ...................CTGCTACAGGACCTGGAGTCCCGGACACCC 14718

    143 LeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAs 159
        :::|||   :::   :::   ||||||      :::      |||::
  14719 TTTCCTTGACAACAGAGAACTAGGAGTAAACCTTCCAGGCAATTCCTAGA 14768

    159 nGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsn 173
        :      :::::::::   |||:::      |||||||||:::
  14769 T...ATCACCTCTGCTCTTGCATTAGTCATTCATCTGACATCA 14808
```

```
seq_name: gb_htg:AC013934

seq_documentation_block:
LOCUS       AC013934              169516 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC013934
VERSION     AC013934.1  GI:6437401
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 169516)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214156 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .169516
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
```

```
BASE COUNT    50796 a  35253 c  34977 g  48490 t
ORIGIN


alignment_scores:
           Quality:   98.50            Length:    221
             Ratio:  0.895              Gaps:     12
 Percent Similarity:  49.774   Percent Identity: 25.792

alignment_block:
 US-09-674-779-2 x AC013934/rev   ..

 Align seg 1/1  to reverse of: AC013934  from: 1  to: 169516

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||||||||   ::::::|||         :::|||:::   ||||||
   90846 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 90802

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                                 :::|||:::   :::|||   |||
   90801 .....CGACGGGATCGTAATGCTGATGGGCATC 90774

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
           :::|||      :::   ::::::   :::   |||:::|||   :::
   90773 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 90730

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
         :::   |||:::      :::   :::   :::|||:::|||
   90729 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 90680

      81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
          |||:::|||               |||      :::|||   ||||||
   90679 TAGGTGGCAG.............CGTCATCGAAGCTATCCTCCACCT. 90646

      98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                        |||:::::::|||:::
   90645 ................................GCAGCGGCGCACGAC 90631

     115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
         |||   ::::::||||||   |||   :::|||   :::   :::
   90630 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 90584

     129 ..........................ProThrLeuHisLeuTyrG 135
                                   |||:::            |
   90583 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 90534

     135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
         ||   :::::::::   :::   ||||||      :::|||   ||||||
   90533 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 90484

     152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
               :::   |||:::      :::   |||   |||   |||:::::
   90483 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 90434

     168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
         :::::|||   |||:::|||:::    |||:::   |||          :::
   90433 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 90384

     185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
               |||   |||   ::::::|||       |||   |||:::
   90383 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 90340

     201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                 :::::::::::::|||   |||       |||   ||||||
   90339 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 90290

     215 yrAlaThrGly 218
           |||   |||
   90289 AGGCGATGGGC 90279

seq_name: gb_in:AC009905

seq_documentation_block:
```

```
LOCUS       AC009905              170662 bp    DNA     linear   INV 22-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 36C-36D, BAC clone
            BACR21E16, complete sequence.
ACCESSION   AC009905
VERSION     AC009905.11  GI:13430983
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 170662)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 36C-36D
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170662)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 22, 2001 this sequence version replaced gi:7264761.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .170662
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="36C-36D"
                     /clone="BACR21E16 (D1090)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    51798 a  32790 c  33582 g  52492 t
ORIGIN


alignment_scores:
           Quality:   98.50            Length:    221
             Ratio:  0.895              Gaps:     12
 Percent Similarity:  49.774   Percent Identity: 25.792

alignment_block:
 US-09-674-779-2 x AC009905   ..
```

```
Align seg 1/1  to: AC009905  from: 1  to: 170662

     17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
        |||||||||   :::::::|||          :::|||:::   ||||||
 165805 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 165849

     33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                                 :::|||:::   :::|||   |||
 165850 .....CGACGGGATCGTAATGCTGATGGGCATC 165877

     50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
          :::|||      :::   ::::::   :::   |||:::|||   :::
 165878 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 165921

     67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
        :::   |||:::      :::   :::   :::|||:::|||
 165922 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 165971

     81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         |||:::|||               |||      :::|||   ||||||
 165972 TAGGTGGCAG.............CGTCATCGAAGCTATCCTCCACCT. 166005

     98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                          |||:::::::|||:::
 166006 ..............................GCAGCGGCGCACGAC 166020

    115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
        |||   ::::::|||||   |||   :::|||   :::   :::
 166021 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 166067

    129 ............................ProThrLeuHisLeuTyrG 135
                                    |||:::            |
 166068 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 166117

    135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
        ||   :::::::::   :::   ||||||      :::|||   ||||||
 166118 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 166167

    152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
             :::   |||:::      :::   |||   |||   |||:::::
 166168 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 166217

    168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
        :::|||   |||:::|||:::    |||:::   |||          :::
 166218 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 166267

    185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
                |||   |||   ::::::|||       |||   |||:::
 166268 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 166311

    201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                ::::::::::::|||   |||       |||   ||||||
 166312 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 166361

    215 yrAlaThrGly 218
           |||   |||
 166362 AGGCGATGGGC 166372

seq_name: gb_in:AC010121

seq_documentation_block:
LOCUS       AC010121              172529 bp    DNA     linear   INV 21-APR-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 36C-36D, BAC clone
            BACR07M13, complete sequence.
ACCESSION   AC010121
VERSION     AC010121.7  GI:13702787
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 172529)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 36C-36D
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172529)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Apr 21, 2001 this sequence version replaced gi:7264762.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES              Location/Qualifiers
     source           1. .172529
                      /organism="Drosophila melanogaster"
                      /strain="y; cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="2L"
                      /map="36C-36D"
                      /clone="BACR07M13 (D1091)"
                      /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                      Drosophila melanogaster BAC library, partial EcoRI in
                      pBACe3.6)"
BASE COUNT    49113 a  35514 c  36013 g  51889 t
ORIGIN


alignment_scores:
            Quality:   98.50              Length:    221
              Ratio:  0.895                Gaps:     12
 Percent Similarity:  49.774   Percent Identity: 25.792

alignment_block:
 US-09-674-779-2 x AC010121   ..

 Align seg 1/1  to: AC010121  from: 1  to: 172529

     17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
        |||||||||   :::::::|||          :::|||:::   ||||||
  52158 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 52202

     33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                                 :::|||:::   :::|||   |||
  52203 .....CGACGGGATCGTAATGCTGATGGGCATC 52230
```

```
    50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
       :::|||      :::   ::::::   :::   |||:::|||   :::
 52231 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 52274

    67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
       :::   |||:::      :::   :::   :::|||:::|||
 52275 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 52324

    81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
        |||:::|||           |||      :::|||   ||||||
 52325 TAGGTGGCAG.............CGTCATCGAAGCTATCCTCCACCT. 52358

    98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                        |||:::::::|||:::
 52359 ................................GCAGCGGCGCACGAC 52373

   115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
       |||   ::::::||||||   |||   :::|||   :::   :::
 52374 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 52420

   129 ..............................ProThrLeuHisLeuTyrG 135
                                       |||:::          |
 52421 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 52470

   135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
       ||   :::::::::   :::   ||||||      :::|||   ||||||
 52471 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 52520

   152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
           :::   |||:::      :::   |||   |||   |||:::::
 52521 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 52570

   168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
       ::::|||   |||:::|||:::   |||:::   |||       :::
 52571 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 52620

   185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
               |||   |||   ::::::|||      |||   |||:::
 52621 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 52664

   201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
             ::::::::::::|||   |||      |||   ||||||
 52665 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 52714

   215 yrAlaThrGly 218
       |||   |||
 52715 AGGCGATGGGC 52725

seq_name: gb_htg:AC087566

seq_documentation_block:
LOCUS       AC087566              231912 bp    DNA     linear   HTG 10-JAN-2001
DEFINITION  Mus musculus clone RP23-127H3, WORKING DRAFT SEQUENCE, 57 unordered
            pieces.
ACCESSION   AC087566
VERSION     AC087566.1  GI:12061432
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 231912)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231912)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_127H3
            -------
            Summary Statistics
            Consensus quality: 150164 bases at least Q40
            Consensus quality: 175115 bases at least Q30
            Consensus quality: 186307 bases at least Q20
            Estimated insert size: 216300; agarose-fp estimation
            Estimated insert size: 226312; sum-of-contigs estimation
            Quality coverage:  3.5 in Q20 bases; agarose-fp estimation
            Quality coverage:  3.34 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 57 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1074: contig of 1074 bp in length
            *     1075     1174: gap of unknown length
            *     1175     2270: contig of 1096 bp in length
            *     2271     2370: gap of unknown length
            *     2371     3511: contig of 1141 bp in length
            *     3512     3611: gap of unknown length
            *     3612     4783: contig of 1172 bp in length
            *     4784     4883: gap of unknown length
            *     4884     6377: contig of 1494 bp in length
            *     6378     6477: gap of unknown length
            *     6478     7711: contig of 1234 bp in length
            *     7712     7811: gap of unknown length
            *     7812     9092: contig of 1281 bp in length
            *     9093     9192: gap of unknown length
            *     9193    10330: contig of 1138 bp in length
            *    10331    10430: gap of unknown length
            *    10431    11677: contig of 1247 bp in length
            *    11678    11777: gap of unknown length
            *    11778    13186: contig of 1409 bp in length
            *    13187    13286: gap of unknown length
            *    13287    14565: contig of 1279 bp in length
            *    14566    14665: gap of unknown length
            *    14666    15787: contig of 1122 bp in length
            *    15788    15887: gap of unknown length
            *    15888    17012: contig of 1125 bp in length
            *    17013    17112: gap of unknown length
            *    17113    18353: contig of 1241 bp in length
            *    18354    18453: gap of unknown length
            *    18454    19611: contig of 1158 bp in length
            *    19612    19711: gap of unknown length
            *    19712    21084: contig of 1373 bp in length
            *    21085    21184: gap of unknown length
            *    21185    22398: contig of 1214 bp in length
            *    22399    22498: gap of unknown length
            *    22499    23629: contig of 1131 bp in length
            *    23630    23729: gap of unknown length
            *    23730    25112: contig of 1383 bp in length
            *    25113    25212: gap of unknown length
            *    25213    26529: contig of 1317 bp in length
            *    26530    26629: gap of unknown length
            *    26630    27773: contig of 1144 bp in length
            *    27774    27873: gap of unknown length
            *    27874    29211: contig of 1338 bp in length
            *    29212    29311: gap of unknown length
            *    29312    30438: contig of 1127 bp in length
            *    30439    30538: gap of unknown length
            *    30539    31930: contig of 1392 bp in length
            *    31931    32030: gap of unknown length
            *    32031    33112: contig of 1082 bp in length
            *    33113    33212: gap of unknown length
            *    33213    34380: contig of 1168 bp in length
```

```
            *    34381    34480: gap of unknown length
            *    34481    35515: contig of 1035 bp in length
            *    35516    35615: gap of unknown length
            *    35616    37145: contig of 1530 bp in length
            *    37146    37245: gap of unknown length
            *    37246    38655: contig of 1410 bp in length
            *    38656    38755: gap of unknown length
            *    38756    40418: contig of 1663 bp in length
            *    40419    40518: gap of unknown length
            *    40519    42089: contig of 1571 bp in length
            *    42090    42189: gap of unknown length
            *    42190    43485: contig of 1296 bp in length
            *    43486    43585: gap of unknown length
            *    43586    45557: contig of 1972 bp in length
            *    45558    45657: gap of unknown length
            *    45658    47561: contig of 1904 bp in length
            *    47562    47661: gap of unknown length
            *    47662    50270: contig of 2609 bp in length
            *    50271    50370: gap of unknown length
            *    50371    52463: contig of 2093 bp in length
            *    52464    52563: gap of unknown length
            *    52564    54241: contig of 1678 bp in length
            *    54242    54341: gap of unknown length
            *    54342    56282: contig of 1941 bp in length
            *    56283    56382: gap of unknown length
            *    56383    57724: contig of 1342 bp in length
            *    57725    57824: gap of unknown length
            *    57825    60404: contig of 2580 bp in length
            *    60405    60504: gap of unknown length
            *    60505    63040: contig of 2536 bp in length
            *    63041    63140: gap of unknown length
            *    63141    64314: contig of 1174 bp in length
            *    64315    64414: gap of unknown length
            *    64415    68063: contig of 3649 bp in length
            *    68064    68163: gap of unknown length
            *    68164    72715: contig of 4552 bp in length
            *    72716    72815: gap of unknown length
            *    72816    79858: contig of 7043 bp in length
            *    79859    79958: gap of unknown length
            *    79959    87305: contig of 7347 bp in length
            *    87306    87405: gap of unknown length
            *    87406    94304: contig of 6899 bp in length
            *    94305    94404: gap of unknown length
            *    94405    99973: contig of 5569 bp in length
            *    99974   100073: gap of unknown length
            *   100074   110741: contig of 10668 bp in length
            *   110742   110841: gap of unknown length
            *   110842   116556: contig of 5715 bp in length
            *   116557   116656: gap of unknown length
            *   116657   126137: contig of 9481 bp in length
            *   126138   126237: gap of unknown length
            *   126238   139661: contig of 13424 bp in length
            *   139662   139761: gap of unknown length
            *   139762   147923: contig of 8162 bp in length
            *   147924   148023: gap of unknown length
            *   148024   160030: contig of 12007 bp in length
            *   160031   160130: gap of unknown length
            *   160131   175441: contig of 15311 bp in length
            *   175442   175541: gap of unknown length
            *   175542   191133: contig of 15592 bp in length
            *   191134   191233: gap of unknown length
            *   191234   231912: contig of 40679 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231912
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-127H3"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    61588 a  52985 c  59076 g  49983 t   8280 others
ORIGIN

alignment_scores:
```

```
           Quality:   98.50            Length:    159
             Ratio:  1.407              Gaps:      7
 Percent Similarity:  44.025   Percent Identity: 27.673

alignment_block:
 US-09-674-779-2 x AC087566   ..

 Align seg 1/1  to: AC087566  from: 1  to: 231912

      49 HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHi 65
         |||||||||   ||||||           |||:::   |||   :::||
  196633 CATACGCACACACATGAT...........CACACAGATTTGTGCACACA 196670

      65 sPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValA 82
         |::::::::::||||||||||:::   :::|||   ::::::::::::|||
  196671 CATGAAATCATGGCTGCAGCTCTTCCAGGCCCACTCCTCATCCCTGGTGA 196720

      82 rgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMet 98
           :::|||         :::|||             :::|||
  196721 TTCATCAGCACTTATGGGAATCTTCCCAAGACTGCGTCATCCCCAGCACC 196770

      99 SerGlnLeu......LeuThrThrAlaArgSerTrpGlnAlaCysGlyHi 113
         ||||||         |||:::||||||         :::   |||
  196771 AGCCAACAGCCTCCCCTCTCCACTGCCTTGTGCTGAAAGGAATGTCAAGA 196820

     113 sGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProT 130
          |||      |||
  196821 TGAGTTAATCCAG..................................... 196833

     130 hrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsn 146
           |||:::   |||                                 |||
  196834 ..CCCCACATGGAACAG.........................AAGAAT 196854

     147 ThrGlnIleArgSerValTyr..............ArgAsnProGluLe 158
          |||   |||:::|||                 |||:::   :::
  196855 AAGCAGACCAGAGAGGTTAAGTGACTTGTTCGTGGCCGCACAGCTAATTA 196904

     158 uAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerA 175
          ||||||   :::      :::      |||             ||||
  196905 GAACCAGATAAGTCCCACCTCTGAAGCTTCT..............TCTG 196939

     175 laIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr 191
         ||:::::      :::|||   |||   :::   ||||||
  196940 CCCTCTCCGGAACGACTCCCAGGCTCACTGTGACATCTCAG......... 196980

     192 GluLeuGlnAsnArgLeuCysGlnTyr 200
                 |||||||||   |||
  196981 ............AGGCTCTGTGGCTAT 196995

seq_name: gb_in:AE003655

seq_documentation_block:
LOCUS       AE003655              253753 bp    DNA     linear   INV 04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386055 section 48
            of 63, complete sequence.
ACCESSION   AE003655 AE002690
VERSION     AE003655.2  GI:10728844
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 253753)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
```

```
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 253753)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7298378.
FEATURES             Location/Qualifiers
     source          1. .253753
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            join(327. .428,1031. .1082,1148. .1482,10017. .10079,
                     15983. .16138,16199. .16414,16480. .16645,16700. .16877,
                     17103. .17168,17962. .18072,20317. .20427,20943. .21009,
                     21072. .22528,22583. .24412,24480. .24610,24678. .24886,
                     24946. .25286)
                     /gene="CLIP-190"
                     /product="CT16052"
                     /db_xref="FLYBASE:FBan0005020"
                     /db_xref="FLYBASE:FBgn0020503"
     gene            <327. .>25286
                     /gene="CLIP-190"
                     /note="CG5020"
                     /map="36C1-36C2"
                     /db_xref="FLYBASE:FBan0005020"
                     /db_xref="FLYBASE:FBgn0020503"
     mRNA            join(520. .1082,1148. .1482,10017. .10079,15983. .16138,
                     16199. .16414,16480. .16645,16700. .16877,17103. .17168,
                     17962. .18072,20317. .20427,20943. .21009,21072. .22528,
                     22583. .24412,24480. .24610,24678. .24886,24946. .25286)
                     /gene="CLIP-190"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT15940"
                     /db_xref="FLYBASE:FBan0005020"
                     /db_xref="FLYBASE:FBgn0020503"
     CDS             join(1218. .1482,10017. .10079,15983. .16138,16199. .16414,
                     16480. .16645,16700. .16877,17103. .17168,17962. .18072,
                     20317. .20427,20943. .21009,21072. .22528,22583. .24412,
                     24480. .24610,24678. .24886,24946. .24992)
                     /gene="CLIP-190"
                     /note="CLIP-190 gene product [alt 1]; Nucleotide sequence
                     of the Celera sequence differs from the published sequence
                     for this transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0005020"
                     /db_xref="FLYBASE:FBgn0020503"
                     /protein_id="AAF53604.1"
                     /db_xref="GI:7298379"
                     /translation="MSDDTSASGGTSAPFPSPVTADPEPGATASKLPGPIRSNIPTPA
                     TSGTGIPQPSKMKAPSSFGSTGSVSKIGRPCCNHTTPKSGPPPREATSMSRESDDNLS
                     SINSAYTDNSSAVLTANTEQFIIGQRVWLGGTRPGQIAFIGDTHFAAGEWAGVVLDEP
                     NGKNDGCVSGKRYFQCEPKRGIFSRLTRLTTYPLAGAQTPTSPLAKSSPDRSRTVSPT
                     ASIRSSMLRSPGIGGKNGMAVGDRVIVSSGFGSRPGILRYLGETQFAPGNWCGVELDE
                     PSGKNDGTVDDIRYFECKPKYGVFVPIAKVSLSPSSKKTRLSRTGSRESLTSIGTMNS
                     IATTATSRMRMNAQQRKSSTPVKPILATPKSQFSMQDLLREKQQHVEKLMVERDLDRE
                     DAQNQALQLQKNINELKARIVELESALDNERKKTEELQCSIDEAQFCGDELNAQSQVY
                     KEKIHDLESKITKLVSATPSLQSILPPDLPSDDGALQEEIAKLQEKMTIQQKEVESRI
                     AEQLEEEQRLRENVKYLNEQIATLQSELVSKDEALEKFSLSECGIENLRRELELLKEE
                     NEKQAQEAQAEFTRKLAEKSVEVLRLSSELQNLKATSDSLESERVNKTDECEILQTEV
                     RMRDEQIRELNQQLDEVTTQLNVQKADSSALDDMLRLQKEGTEEKSTLLEKTEKELVQ
                     SKEQAAKTLNDKEQLEKQISDLKQLAEQEKLVREMTENAINQIQLEKESIEQQLALKQ
                     NELEDFQKKQSESEVHLQEIKAQNTQKDFELVESGESLKKLQQQLEQKTLGHEKLQAA
                     LEELKKEKETIIKEKEQELQQLQSKSAESESALKVVQVQLEQLQQQAAASGEEGSKTV
                     AKLHDEISQLKSQAEETQSELKSTQSNLEAKSKQLEAANGSLEEEAKKSGHLLEQITK
                     LKSEVGETQAALSSCHTDVESKTKQLEAANAALEKVNKEYAESRAEASDLQDKVKEIT
                     DTLHAELQAERSSSSALHTKLSKFSDEIATGHKELTSKADAWSQEMLQKEKELQELRQ
                     QLQDSQDSQTKLKAEGERKEKSFEESIKNLQEEVTKAKTENLELSTGTQTTIKDLQER
                     LEITNAELQHKEKMASEDAQKIADLKTLVEAIQVANANISATNAELSTVLEVLQAEKS
                     ETNHIFELFEMEADMNSERLIEKVTGIKEELKETHLQLDERQKKFEELEEKLKQAQQS
                     EQKLQQESQTSKEKLTEIQQSLQELQDSVKQKEELVQNLEEKVRESSSIIEAQNTKLN
                     ESNVQLENKTSCLKETQDQLLESQKKEKQLQEEAAKLSGELQQVQEANGDIKDSLVKV
                     EELVKVLEEKLQAATSQLDAQQATNKELQELLVKSQENEGNLQGESLAVTEKLQQLEQ
CDS                  join(1218. .1482,10017. .10079,15983. .16138,16199. .16414,
                     16480. .16645,16700. .16877,17103. .17168,17962. .18072,
                     20317. .20427,20943. .21009,21072. .22528,22583. .24412,
                     24480. .24610,24678. .24886,24946. .24992)
                     /gene="CLIP-190"
                     /note="CLIP-190 gene product [alt 2]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0005020"
                     /db_xref="FLYBASE:FBgn0020503"
                     /protein_id="AAF53605.1"
                     /db_xref="GI:7298380"
                     /translation="MSDDTSASGGTSAPFPSPVTADPEPGATASKLPGPIRSNIPTPA
                     TSGTGIPQPSKMKAPSSFGSTGSVSKIGRPCCNHTTPKSGPPPREATSMSRESDDNLS
                     SINSAYTDNSSAVLTANTEQFIIGQRVWLGGTRPGQIAFIGDTHFAAGEWAGVVLDEP
                     NGKNDGCVSGKRYFQCEPKRGIFSRLTRLTTYPLAGAQTPTSPLAKSSPDRSRTVSPT
                     ASIRSSMLRSPGIGGKNGMAVGDRVIVSSGFGSRPGILRYLGETQFAPGNWCGVELDE
                     PSGKNDGTVDDIRYFECKPKYGVFVPIAKVSLSPSSKKTRLSRTGSRESLTSIGTMNS
                     IATTATSRMRMNAQQRKSSTPVKPILATPKSQFSMQDLLREKQQHVEKLMVERDLDRE
                     DAQNQALQLQKNINELKARIVELESALDNERKKTEELQCSIDEAQFCGDELNAQSQVY
                     KEKIHDLESKITKLVSATPSLQSILPPDLPSDDGALQEEIAKLQEKMTIQQKEVESRI
                     AEQLEEEQRLRENVKYLNEQIATLQSELVSKDEALEKFSLSECGIENLRRELELLKEE
                     NEKQAQEAQAEFTRKLAEKSVEVLRLSSELQNLKATSDSLESERVNKTDECEILQTEV
                     RMRDEQIRELNQQLDEVTTQLNVQKADSSALDDMLRLQKEGTEEKSTLLEKTEKELVQ
                     SKEQAAKTLNDKEQLEKQISDLKQLAEQEKLVREMTENAINQIQLEKESIEQQLALKQ
                     NELEDFQKKQSESEVHLQEIKAQNTQKDFELVESGESLKKLQQQLEQKTLGHEKLQAA
                     LEELKKEKETIIKEKEQELQQLQSKSAESESALKVVQVQLEQLQQQAAASGEEGSKTV
                     AKLHDEISQLKSQAEETQSELKSTQSNLEAKSKQLEAANGSLEEEAKKSGHLLEQITK
                     LKSEVGETQAALSSCHTDVESKTKQLEAANAALEKVNKEYAESRAEASDLQDKVKEIT
                     DTLHAELQAERSSSSALHTKLSKFSDEIATGHKELTSKADAWSQEMLQKEKELQELRQ
                     QLQDSQDSQTKLKAEGERKEKSFEESIKNLQEEVTKAKTENLELSTGTQTTIKDLQER
                     LEITNAELQHKEKMASEDAQKIADLKTLVEAIQVANANISATNAELSTVLEVLQAEKS
                     ETNHIFELFEMEADMNSERLIEKVTGIKEELKETHLQLDERQKKFEELEEKLKQAQQS
                     EQKLQQESQTSKEKLTEIQQSLQELQDSVKQKEELVQNLEEKVRESSSIIEAQNTKLN
                     ESNVQLENKTSCLKETQDQLLESQKKEKQLQEEAAKLSGELQQVQEANGDIKDSLVKV
                     EELVKVLEEKLQAATSQLDAQQATNKELQELLVKSQENEGNLQGESLAVTEKLQQLEQ
```

```
                     ANGELKEALCQKENGLKELQGKLDESNTVLESQKKSHNEIQDKLEQAQQKERTLQEET
                     SKLAEQLSQLKQANEELQKSLQQKQLLLEKGNEFDTQLAEYQKVIDEMDDAASVKSAL
                     LEQLQNRVAELETALRQANDAQKTAYLETKELRRQLESLELEKSREVLSLKAQMNGAS
                     SRSGKGDEVESLDIETSLAKINFLNSIIADMQQKNDALKAKVQTLETLPMDFTKPHAF
                     DALTKRKPAPRLFCDICDEFDQHDTEDCPIQGSEDQDYSTPSSESNNNEKERKLPAPR
                     KYCDSCEVFGHDTSECADDETY"
     mRNA            complement(join(<25345. .25380,25443. .25617,25675. .25764,
                     25839. .>25891))
                     /gene="CG6840"
                     /product="CT21109"
                     /db_xref="FLYBASE:FBan0006840"
                     /db_xref="FLYBASE:FBgn0032634"
     gene            complement(<25345. .>25891)
                     /gene="CG6840"
                     /map="36C1-36C2"
                     /db_xref="FLYBASE:FBan0006840"
                     /db_xref="FLYBASE:FBgn0032634"
     CDS             complement(join(25345. .25380,25443. .25617,25675. .25764,
                     25839. .25891))
                     /gene="CG6840"
                     /note="CG6840 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0006840"
                     /db_xref="FLYBASE:FBgn0032634"
                     /protein_id="AAF53606.1"
                     /db_xref="GI:7298381"
                     /translation="MNAPPTFESFLLYEGEKKIIKELDTKVTNAAIFTINKEDHTLGN
                     MIRNQLLKDPNVLFAGYKVPHPLEHKFVIRIQTTADYSPQEAFMNAITDLLAELSLFE
                     ERFKDAIKEKKEGGD"
     mRNA            join(26244. .27205,28382. .>28758)
                     /gene="CG15141"
                     /product="CT35038"
                     /db_xref="FLYBASE:FBan0015141"

alignment_scores:
           Quality:   98.50            Length:    221
             Ratio:  0.895              Gaps:     12
 Percent Similarity:  49.774   Percent Identity: 25.792

alignment_block:
 US-09-674-779-2 x AE003655   ..

 Align seg 1/1  to: AE003655  from: 1  to: 253753

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         ||||||||   :::::::|||         :::|||:::   ||||||
   56335 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 56379

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                               :::|||:::   :::|||   |||
   56380 .....CGACGGGATCGTAATGCTGATGGGCATC 56407

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
           :::|||      :::   ::::::   :::   |||:::|||   :::
   56408 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 56451

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
         :::   |||:::      :::   :::   :::|||:::|||
   56452 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 56501

      81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
          |||:::|||                |||      :::|||   ||||||
   56502 TAGGTGGCAG.............CGTCATCGAAGCTATCCTCCACCT. 56535

      98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                        |||:::::::|||:::
   56536 .................................GCAGCGGCGCACGAC 56550

     115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
         |||   :::::::|||:||   |||   :::|||   :::   :::
   56551 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 56597

     129 ............................ProThrLeuHisLeuTyrG 135
                                          |||:::           |
   56598 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 56647

     135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
         ||   :::::::::   :::   ||||||      :::|||   ||||||
   56648 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 56697

     152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
               :::   |||:::      :::   |||   |||   |||:::::
   56698 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 56747

     168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
         ::::|||   |||:::|||:::    |||:::   |||          :::
   56748 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 56797

     185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
                 |||   |||   :::::::|||      |||   |||:::
   56798 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 56841

     201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                :::::::::::::|||   |||      |||   ||||||
   56842 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 56891

     215 yrAlaThrGly 218
           |||   |||
   56892 AGGCGATGGGC 56902

seq_name: gb_pr:HS278N12

seq_documentation_block:
LOCUS       HS278N12               78533 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Human DNA sequence from clone 278N12 on chromosome 6q25.1-26.
            Contains ESTs, an STS, GSSs and genonic marker D6S425, complete
            sequence.
ACCESSION   AL033376
VERSION     AL033376.17  GI:5360990
KEYWORDS    HTG; D6S425.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 78533)
  AUTHORS   Mashreghi-Mohammadi,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 6, 1999 this sequence version replaced gi:5262828.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL
            This sequence is the entire insert of clone 278N12. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            278N12 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .78533
```

```
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /map="q25.1-26"
                /clone="RP1-278N12"
                /clone_lib="RPCI-1"
repeat_region   1. .121
                /note="L1MA8 repeat: matches 5913. .6034 of consensus"
repeat_region   122. .412
                /note="AluSg repeat: matches 1. .290 of consensus"
repeat_region   413. .669
                /note="L1MA8 repeat: matches 6034. .6286 of consensus"
repeat_region   706. .853
                /note="AluSg/x repeat: matches 135. .299 of consensus"
repeat_region   857. .1167
                /note="AluSx repeat: matches 1. .308 of consensus"
repeat_region   1501. .1812
                /note="AluJo repeat: matches 3. .310 of consensus"
repeat_region   2652. .2949
                /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region   2972. .3082
                /note="L2 repeat: matches 2572. .2697 of consensus"
repeat_region   3620. .3924
                /note="AluSg repeat: matches 1. .273 of consensus"
repeat_region   4005. .4175
                /note="TIGGER1 repeat: matches 2238. .2416 of consensus"
repeat_region   4176. .4659
                /note="L1MA8 repeat: matches 5784. .6285 of consensus"
repeat_region   4685. .5034
                /note="L1ME3 repeat: matches 4910. .5235 of consensus"
repeat_region   5035. .5342
                /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region   5343. .6094
                /note="L1ME3 repeat: matches 5235. .6139 of consensus"
repeat_region   6335. .6508
                /note="AluJb repeat: matches 134. .311 of consensus"
repeat_region   8493. .8614
                /note="FLAM_A repeat: matches 1. .122 of consensus"
repeat_region   8616. .8691
                /note="38 copies 2 mer ta 71% conserved"
repeat_region   8700. .8997
                /note="AluJo repeat: matches 1. .300 of consensus"
repeat_region   9263. .9573
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   9584. .10001
                /note="L1M4 repeat: matches 3585. .4012 of consensus"
repeat_region   10003. .10136
                /note="AluJb repeat: matches 2. .135 of consensus"
misc_feature    10090. .10460
                /note="match: STS Z23751; genomic marker D6S425"
repeat_region   10139. .10168
                /note="15 copies 2 mer ac 97% conserved"
repeat_region   10170. .10307
                /note="L1M4 repeat: matches 4063. .4203 of consensus"
repeat_region   10323. .10616
                /note="AluSq repeat: matches 1. .296 of consensus"
repeat_region   10625. .10812
                /note="L1MC5 repeat: matches 7481. .7675 of consensus"
misc_feature    complement(10716. .11105)
                /note="match: GSS AQ218681"
repeat_region   10931. .11376
                /note="LTR16C repeat: matches 1. .387 of consensus"
repeat_region   11377. .11439
                /note="L1MC5 repeat: matches 7718. .7779 of consensus"
repeat_region   11440. .11725
                /note="AluSq repeat: matches 3. .288 of consensus"
repeat_region   11726. .11855
                /note="L1MC5 repeat: matches 7779. .7913 of consensus"
repeat_region   12052. .12192
                /note="FLAM_C repeat: matches 2. .138 of consensus"
repeat_region   12293. .12590
                /note="AluSx repeat: matches 1. .298 of consensus"
repeat_region   12608. .12943
                /note="L1M4 repeat: matches 5125. .5467 of consensus"
misc_feature    12820. .13220
                /note="match: GSS B80534"
repeat_region   13039. .13120
                /note="41 copies 2 mer ta 73% conserved"
repeat_region   13312. .13449
                /note="L2 repeat: matches 2612. .2745 of consensus"
repeat_region   13711. .14015
                /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region   14053. .15990
                /note="L1PA14 repeat: matches 4207. .6144 of consensus"
repeat_region   16225. .16366
                /note="FLAM_C repeat: matches 1. .133 of consensus"
repeat_region   16898. .17018
                /note="FLAM_C repeat: matches 2. .122 of consensus"
repeat_region   17081. .17383
                /note="AluJo repeat: matches 1. .308 of consensus"
repeat_region   17384. .17421
                /note="19 copies 2 mer ga 82% conserved"
misc_feature    17669. .17925
                /note="match: GSS AQ527764"
repeat_region   18093. .18388
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   18669. .18790
                /note="MER50 repeat: matches 1. .129 of consensus"
repeat_region   18793. .19028
                /note="AluSx repeat: matches 42. .276 of consensus"
repeat_region   19030. .19271
                /note="121 copies 2 mer aa 65% conserved"
repeat_region   19309. .19898
                /note="MER50 repeat: matches 117. .711 of consensus"
misc_feature    complement(20062. .20712)
                /note="match: ESTs AI269581 AA226135 AI476542 AA226134"
repeat_region   20064. .20162
                /note="FLAM_C repeat: matches 1. .133 of consensus"
misc_feature    20399. .20867
                /note="match: GSS AQ499824"
repeat_region   20934. .21330
                /note="MLT2B repeat: matches 1. .404 of consensus"
repeat_region   21333. .21462
                /note="65 copies 2 mer at 76% conserved"
repeat_region   21464. .21599
                /note="68 copies 2 mer at 76% conserved"
repeat_region   22069. .22142
                /note="MIR repeat: matches 68. .143 of consensus"
repeat_region   22451. .22530
                /note="MIR repeat: matches 52. .131 of consensus"
repeat_region   22583. .23228
                /note="LTR29 repeat: matches 5. .619 of consensus"
repeat_region   23254. .23320
                /note="L1MA4A repeat: matches 5149. .5211 of consensus"
repeat_region   23321. .23448
                /note="FLAM_C repeat: matches 1. .127 of consensus"
repeat_region   23449. .23462
                /note="L1MA4A repeat: matches 5211. .5224 of consensus"
repeat_region   23461. .24364
                /note="L1MA4A repeat: matches 5191. .6045 of consensus"
repeat_region   24365. .24670
                /note="AluJb repeat: matches 1. .305 of consensus"
repeat_region   24671. .24917
                /note="L1MA4A repeat: matches 6045. .6300 of consensus"
misc_feature    complement(25321. .25698)
                /note="match: EST H75280"
misc_feature    complement(25359. .25909)
                /note="match: GSS AQ592624"
repeat_region   25533. .25673
                /note="LTR23 repeat: matches 1. .146 of consensus"
repeat_region   25717. .25796
                /note="MIR repeat: matches 52. .131 of consensus"
repeat_region   26104. .26177
                /note="MIR repeat: matches 68. .143 of consensus"
repeat_region   26431. .26726
                /note="AluJb repeat: matches 27. .304 of consensus"
```

```
     misc_feature    complement(26669. .27116)
                     /note="match: GSS AQ138816"
     repeat_region   26727. .27005
                     /note="LTR23 repeat: matches 135. .489 of consensus"
     repeat_region   27207. .27338
                     /note="FLAM_C repeat: matches 2. .133 of consensus"
     repeat_region   27347. .27663
                     /note="AluJo repeat: matches 1. .312 of consensus"
     repeat_region   27855. .28003
                     /note="MER4B repeat: matches 481. .610 of consensus"
     repeat_region   28004. .28318

alignment_scores:
            Quality:   98.00            Length:    271
              Ratio:  0.803              Gaps:     15
 Percent Similarity:  45.018   Percent Identity: 22.509

alignment_block:
 US-09-674-779-2 x HS278N12   ..

 Align seg 1/1  to: HS278N12  from: 1  to: 78533

      22 AlaProIleProThrAsnProGlnValSerProIleLysThrProSerVa 38
         :::|||:::|||:::   ||||||   ||||||         ||||||::
   42082 TCACCTCTACCATCACTTCCTCAACCATCTCCACCATGTCTACCATCACT 42131

      38 lLeu......IleThrLysAspLysIleGlyAspHisHisThrHisGluH 53
         :|||      :::|||:::         ::::::||||||   |||:::|
   42132 TCTACCATCTGTCACCTCCACCGTCACCTCCAACCACCATCTCCACCACC 42181

      53 isAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrp 69
         ||   :::         |||      |||:::   |||::::::
   42182 ACCTCCACCACCACCTCCACCATCTCCTCCACCACCACCTCCACCATCTC 42231

      70 Leu......GlnMetHisHisAla........................... 75
         |||      ::::::||||||
   42232 CTCCACAACCACCTCCACCATCTCTACCGTCACTTCCACCATCACCTCTA 42281

      76 ....ThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgL 91
             ::::::::::::      :::   :::   :::|||:::   ||||
   42282 CCACTCCAATCACCATTTCTACCATCTCCATTACCACCTCCACCATCGCC 42331

      91 euGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSer 107
         ||   |||:::|||      :::   :::
   42332 TCCACAACCACCTCTACCATCTCCACCGT..................... 42360

     108 TrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGlu........ 121
                     |||      ::::::||||||   :::
   42361 ..............CACTTCCACCATCACCTCCCCTACGACCACCACTT 42395

     122 ............................HisLeuTrpGlyGlnIleValP 129
                                     ||||||      ::::::   |
   42396 CCAACACCACCCCCTTTATCTCCACCTCCACCTCCACCACCATGTATCGC 42445

     129 roThrLeuHisLeuTyr..............GlnAspLeuLysSerArg 140
         ||||||||   ||||||               :::|||   :::|||
   42446 CAACCCTTTCACTTTACTGACCTGGGGATTTTTCTGATCCCCGTTCTACA 42495

     141 Gly..........................IleLeuProAlaAsnThrGl 148
                                      :::|||   :::::::|||
   42496 CACCCTTTCTTTACTCCACACCTTCATATTCTCCTCTCCTCCTCCACTCT 42545

     148 nIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGly. 164
          |||         :::   |||   ::::::   :::|||   ::::::
   42546 GATC.........TTTTCTAACATTCACTTTTTAAAGTGCCTGAGTAACA 42586

     165 .....AlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
                |||||||:::                              :::
   42587 GAGATAAAGCAATGGCT..............................TTC 42606

     180 ValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnAr 196
         :::|||                  :::::::::||||||
   42607 CTACCA.................CATTCATTTTATGAA........... 42627

     196 gLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuG 213
          |||||||||   ::::::   |||:::   |||
   42628 .TTATGCCAGTTATTCATT...CATTCAACAAAT................ 42657

     213 lyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLys 229
                              |||:::   :::         |||
   42658 ....................ATTTACTGAGAGCAGGCACTGTTCTGAGCA 42687

     230 TrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuPr 246
         |||:::      ::::::::::::|||||||||||   ::::::
   42688 CTTGGGAGC......ACAAGTAACTCCTCCATTTGCAGGATTCTCTTTAA 42731

     246 oLysAsnLysLeu 250
         :::||||||
   42732 GGCTAAAAAGCTT 42744

seq_name: gb_htg:AC097543

seq_documentation_block:
LOCUS       AC097543              146321 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-126I1, *** SEQUENCING IN PROGRESS
            ***, 74 unordered pieces.
ACCESSION   AC097543
VERSION     AC097543.4  GI:17973388
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 146321)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 146321)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17064342.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHUY
                Center clone name: CH230-126I1
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 118791 bases at least Q40
                Consensus quality: 127201 bases at least Q30
                Consensus quality: 132476 bases at least Q20
                Estimated insert size: 94009; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 1.3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 74 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5167: contig of 5167 bp in length
            *     5168     5267: gap of unknown length
            *     5268     8729: contig of 3462 bp in length
            *     8730     8829: gap of unknown length
            *     8830    12812: contig of 3983 bp in length
            *    12813    12912: gap of unknown length
            *    12913    16451: contig of 3539 bp in length
            *    16452    16551: gap of unknown length
            *    16552    19971: contig of 3420 bp in length
            *    19972    20071: gap of unknown length
            *    20072    23215: contig of 3144 bp in length
            *    23216    23315: gap of unknown length
            *    23316    27807: contig of 4492 bp in length
            *    27808    27907: gap of unknown length
            *    27908    30577: contig of 2670 bp in length
            *    30578    30677: gap of unknown length
            *    30678    34125: contig of 3448 bp in length
            *    34126    34225: gap of unknown length
            *    34226    36057: contig of 1832 bp in length
            *    36058    36157: gap of unknown length
            *    36158    38083: contig of 1926 bp in length
            *    38084    38183: gap of unknown length
            *    38184    40595: contig of 2412 bp in length
            *    40596    40695: gap of unknown length
            *    40696    41732: contig of 1037 bp in length
            *    41733    41832: gap of unknown length
            *    41833    44795: contig of 2963 bp in length
            *    44796    44895: gap of unknown length
            *    44896    47736: contig of 2841 bp in length
            *    47737    47836: gap of unknown length
            *    47837    49547: contig of 1711 bp in length
            *    49548    49647: gap of unknown length
            *    49648    51800: contig of 2153 bp in length
            *    51801    51900: gap of unknown length
            *    51901    54832: contig of 2932 bp in length
            *    54833    54932: gap of unknown length
            *    54933    57421: contig of 2489 bp in length
            *    57422    57521: gap of unknown length
            *    57522    59641: contig of 2120 bp in length
            *    59642    59741: gap of unknown length
            *    59742    61775: contig of 2034 bp in length
            *    61776    61875: gap of unknown length
            *    61876    63221: contig of 1346 bp in length
            *    63222    63321: gap of unknown length
            *    63322    64796: contig of 1475 bp in length
            *    64797    64896: gap of unknown length
            *    64897    66314: contig of 1418 bp in length
            *    66315    66414: gap of unknown length
            *    66415    68369: contig of 1955 bp in length
            *    68370    68469: gap of unknown length
            *    68470    70263: contig of 1794 bp in length
            *    70264    70363: gap of unknown length
            *    70364    72846: contig of 2483 bp in length
            *    72847    72946: gap of unknown length
            *    72947    75430: contig of 2484 bp in length
            *    75431    75530: gap of unknown length
            *    75531    77680: contig of 2150 bp in length
            *    77681    77780: gap of unknown length
            *    77781    79155: contig of 1375 bp in length
            *    79156    79255: gap of unknown length
            *    79256    81495: contig of 2240 bp in length
            *    81496    81595: gap of unknown length
            *    81596    83077: contig of 1482 bp in length
            *    83078    83177: gap of unknown length
            *    83178    84399: contig of 1222 bp in length
            *    84400    84499: gap of unknown length
            *    84500    85780: contig of 1281 bp in length
            *    85781    85880: gap of unknown length
            *    85881    87000: contig of 1120 bp in length
            *    87001    87100: gap of unknown length
            *    87101    88656: contig of 1556 bp in length
            *    88657    88756: gap of unknown length
            *    88757    89848: contig of 1092 bp in length
            *    89849    89948: gap of unknown length
            *    89949    92040: contig of 2092 bp in length
            *    92041    92140: gap of unknown length
            *    92141    94019: contig of 1879 bp in length
            *    94020    94119: gap of unknown length
            *    94120    95784: contig of 1665 bp in length
            *    95785    95884: gap of unknown length
            *    95885    97071: contig of 1187 bp in length
            *    97072    97171: gap of unknown length
            *    97172    98387: contig of 1216 bp in length
            *    98388    98487: gap of unknown length
            *    98488    99734: contig of 1247 bp in length
            *    99735    99834: gap of unknown length
            *    99835   101634: contig of 1800 bp in length
            *   101635   101734: gap of unknown length
            *   101735   102787: contig of 1053 bp in length
            *   102788   102887: gap of unknown length
            *   102888   104247: contig of 1360 bp in length
            *   104248   104347: gap of unknown length
            *   104348   105668: contig of 1321 bp in length
            *   105669   105768: gap of unknown length
            *   105769   107032: contig of 1264 bp in length
            *   107033   107132: gap of unknown length
            *   107133   108500: contig of 1368 bp in length
            *   108501   108600: gap of unknown length
            *   108601   110253: contig of 1653 bp in length
            *   110254   110353: gap of unknown length
            *   110354   112309: contig of 1956 bp in length
            *   112310   112409: gap of unknown length
            *   112410   113605: contig of 1196 bp in length
            *   113606   113705: gap of unknown length
            *   113706   114730: contig of 1025 bp in length
            *   114731   114830: gap of unknown length
            *   114831   116310: contig of 1480 bp in length
            *   116311   116410: gap of unknown length
            *   116411   117540: contig of 1130 bp in length
            *   117541   117640: gap of unknown length
            *   117641   118994: contig of 1354 bp in length
```

```
          *   118995   119094: gap of unknown length

alignment_scores:
            Quality:   98.00            Length:    188
              Ratio:  1.289              Gaps:     10
 Percent Similarity:  40.426   Percent Identity: 20.745

alignment_block:
 US-09-674-779-2 x AC097543   ..

 Align seg 1/1  to: AC097543  from: 1  to: 146321

      20 CysSerAlaPro.........IleProThrAsnProGlnValSerProIl 33
         ||||||:::            :::   ::::::||||||:::   |||::
  110866 TGCTCTTCTCAAACAGGTTGTCTTGGGTCTCAGCCTCAGTTACCTCCACT 110915

      33 eLys.........ThrProSerValLeuIleThrLysAspLysIleGly. 46
         :|||         :::||||||   ::::::||||||   :::||||||
  110916 GAAAATGGGTACATCTCCCTCTAGGGTTGTTACAAAGATCTCGATAGGTT 110965

      47 ...............................AspHisHisThrHisGlu 52
                                                 |||||||||
  110966 ATGTGTTGACAGTTCACACAGAGACACACAGACACACACACACACACA 111015

      53 HisAspGluSerValSerHisVal........................... 60
         |||   ::::::   :::|||:::
  111016 CACACACACACAAAGACCCATATACACAGAGGGACACACACAAACACACA 111065

      61 .............................................GlyL 62
                                                      :
  111066 TACAGATATATAGATATACACGTAGGCATGCCCACAGAGACAGTCACACA 111115

      62 euGlnAlaHisPheGluThrTrpLeuGlnMetHis.............. 73
         ::|||||||||   :::|||      :::||||||
  111116 TACAGGCACACACACAGACACATGCACACATGCACACACGCGCACACACA 111165

      74 HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerAr 90
         ||||||:::            :::|||   :::|||:::   ::::::::
  111166 CATGCGAGC............GCGCGCGCACACGCACACACACACACACA 111203

      90 gLeuGly............AsnTyrLeuProProMetSerGlnLeuLeuT 103
         :   |||            :::      ||||||::::::::   :::
  111204 CACGGGCTGGAACCCACACAGCTGCAGCCCTCCCCTCACAGAGACGATTC 111253

     103 hrThrAlaArgSerTrp.............................. 108
           :::   :::   |||
  111254 TTTCCCACCACCCCTGGGAAATGATCCTTTCCTCACGGTCGTGACTCAGG 111303

     109 ................................GlnAlaCysGlyHi 113
                                        |||:::|||
  111304 GATAGTCTCTCCCCTTTCCTTGGTTTTCTATCCACACAGGGATGT..... 111348

     113 sGluProTyrGlnLeuProPro......GluHisLeuTrpGlyGlnIleV 128
                    :::|||||||      :::      ||||||::::::
  111349 ..........CTAGTTCCTCCTTAGAGTTCTGCATAGTGGGGAAAACTGG 111388

     128 alProThrLeuHis 132
           |||:::   |||
  111389 GACCTGTGAGGCAC 111402

seq_name: gb_in:PMHISH2A

seq_documentation_block:
LOCUS       PMHISH2A                 790 bp    DNA     linear   INV 24-OCT-2001
DEFINITION  Psammechinus miliaris gene for histone H2A, clone h19).
ACCESSION   X01346
VERSION     X01346.1  GI:10030
KEYWORDS    histone; histone H2A.
SOURCE      sand urchin.
  ORGANISM  Psammechinus miliaris
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;
            Psammechinus.

REFERENCE   1  (bases 1 to 790)
  AUTHORS   Busslinger,M., Portmann,R., Irminger,J.C. and Birnstiel,M.L.
  TITLE     Ubiquitous and gene-specific regulatory 5' sequences in a sea
            urchin histone DNA clone coding for histone protein variants
  JOURNAL   Nucleic Acids Res. 8 (5), 957-977 (1980)
  MEDLINE   81076674
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Psammechinus miliaris"
                     /db_xref="taxon:7660"
     misc_feature    298. .302
                     /note="conserved sequence"
     -10_signal      312. .321
     misc_feature    344. .350
                     /note="putative cap sequence,transcription initiation
                     region"
     misc_feature    412. .418
                     /note="conserved sequence"
     CDS             416. .790
                     /codon_start=1
                     /product="histone H2A"
                     /protein_id="CAA25633.1"
                     /db_xref="GI:10031"
                     /db_xref="SWISS-PROT:P02271"
                     /translation="MSGRGKSGKARTKAKTRSSRAGLQFPVGRVHRFLRKGNYAKRVG
                     GGAPVYMAAVLEYLTAEILELAGNAARDNKKSRIIPRHLQLAVRNDEELNKLLGGVTI
                     AQGGVLPNIQAVLLPKKTAKSS"
BASE COUNT      220 a    198 c    186 g    186 t
ORIGIN


alignment_scores:
            Quality:   97.50            Length:    155
              Ratio:  1.121              Gaps:      9
 Percent Similarity:  56.129   Percent Identity: 26.452

alignment_block:
 US-09-674-779-2 x PMHISH2A   ..

 Align seg 1/1  to: PMHISH2A  from: 1  to: 790

      28 ProGlnValSerProIleLysThrProSerVal.LeuIleThrLysAspL 44
         |||   |||:::   |||   |||||||:::::: :::   ||||||
     284 CCGACGGTCACTGCGATCCTAACCCCAGGTATAAATGGCCACCAAAACGC 333

      44 ysIleGlyAspHisHisThrHis.GluHisAspGluSerValSerHisVa 60
           :::|||   |||   :::||| ||||||          ::::::::::
     334 TGCTGGGCATCCATTCAAGTCATCGAACACTGTTACGTTCTGAACTACGC 383

      60 lGly..............LeuGlnAlaHisPheGluThrTrpLeuGlnM 72
         ::::                |||   |||   :::::::|||   :::
     384 TTCCGATTTATTCTAAACTCATCAAAAACAT...CATGTCTGGCAGAGGA 430

      72 etHis............HisAlaThrLysGlnGluValValArgTyrGln 84
           :::            |||      |||:::   ::::::      :::
     431 AAGAGTGGAAAGGCCCGCACCAAGGCAAAGACGCGCTCATCCCGTGCAGG 480

      85 AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLe 101
         |||   ::::::|||          :::   :::   |||   :::||||||
     481 GCTCCAGTTTCCAGTGGGACGTGTTCATCGGTTTCTCCGAAAGGGCAACT 530

     101 uLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnL 118
         |          :::|||   |||:::::::|||   |||
     531 ATGCAAAGAGGGTCGGCGGTGGAGCTCCTGTCTACAT.........GGCT 571

     118 euProPro.GluHisLeuTrpGlyGlnIleValProThrLeu...HisLe 133
           ||| |||:::|||   ::::::|||:::          :::
     572 GCCGTCCTAGAGTACCTCACTGCCGAAATCTTGGAACTCGCAGGCAACGC 621

     133 uTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleA 150
            :::|||   |||:::      |||:::|||   :::   |||:::
     622 TGCCCGCGACAACAAGAAATCTAGGATCATCCCACGCCACCTTCAACTC. 670
```

```
     150 rgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAla 166
           :::|||   ::::::   |||||||||:::      ||||||::::::
     671 ..GCTGTGCGTAATGATGAAGAACTCAACAAGCTTTTGGGTGGGGTGACG 718

     167 MetSerLys 169
         :::::::::
     719 ATCGCTCAA 727

seq_name: gb_in:SUPHISB4

seq_documentation_block:
LOCUS       SUPHISB4                872 bp    DNA     linear   INV 17-MAR-1994
DEFINITION  P.miliaris histone H2A gene, clone h19, complete cds, clone h19.
ACCESSION   M10559 M10550 M10561 M10646
VERSION     M10559.1  GI:161382
KEYWORDS    histone; histone H2A.
SEGMENT     4 of 5
SOURCE      Psammechinus miliaris sperm DNA.
  ORGANISM  Psammechinus miliaris
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;
            Psammechinus.
REFERENCE   1  (bases 788 to 872)
  AUTHORS   Busslinger,M., Portmann,R. and Birnstiel,M.L.
  TITLE     A regulatory sequence near the 3' end of sea urchin histone genes
  JOURNAL   Nucleic Acids Res. 6, 2997-3008 (1979)
  MEDLINE   80034444
REFERENCE   2  (bases 1 to 790)
  AUTHORS   Busslinger,M., Portmann,R., Irminger,J.C. and Birnstiel,M.L.
  TITLE     Ubiquitous and gene-specific regulatory 5' sequences in a sea
            urchin histone DNA clone coding for histone protein variants
  JOURNAL   Nucleic Acids Res. 8, 957-977 (1980)
  MEDLINE   81076674
COMMENT     Multiple accession numbers refer to duplicate data from [2].
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Psammechinus miliaris"
                     /db_xref="taxon:7660"
                     /tissue_type="sperm"
     CDS             416. .790
                     /codon_start=1
                     /product="histone H2A"
                     /protein_id="AAA30027.1"
                     /db_xref="GI:161388"
                     /translation="MSGRGKSGKARTKAKTRSSRAGLQFPVGRVHRFLRKGNYAKRVG
                     GGAPVYMAAVLEYLTAEILELAGNAARDNKKSRIIPRHLQLAVRNDEELNKLLGGVTI
                     AQGGVLPNIQAVLLPKKTAKSS"
BASE COUNT      243 a    220 c    200 g    209 t
ORIGIN      About 40 bp after segment 3.


alignment_scores:
            Quality:   97.50            Length:    155
              Ratio:  1.121              Gaps:      9
 Percent Similarity:  56.129   Percent Identity: 26.452

alignment_block:
 US-09-674-779-2 x SUPHISB4   ..

 Align seg 1/1  to: SUPHISB4  from: 1  to: 872

      28 ProGlnValSerProIleLysThrProSerVal.LeuIleThrLysAspL 44
         |||   |||:::   |||   ||||||:::::: :::   ||||||
     284 CCGACGGTCACTGCGATCCTAACCCCAGGTATAAATGGCCACCAAAACGC 333

      44 ysIleGlyAspHisHisThrHis.GluHisAspGluSerValSerHisVa 60
         :::|||   |||   :::||| ||||||           ::::::::::::
     334 TGCTGGGCATCCATTCAAGTCATCGAACACTGTTACGTTCTGAACTACGC 383

      60 lGly..............LeuGlnAlaHisPheGluThrTrpLeuGlnM 72
         ::::                  |||   |||   :::::::|||   :::
     384 TTCCGATTTATTCTAAACTCATCAAAAACAT...CATGTCTGGCAGAGGA 430

      72 etHis............HisAlaThrLysGlnGluValValArgTyrGln 84
           :::            |||      |||:::   ::::::      :::
     431 AAGAGTGGAAAGGCCCGCACCAAGGCAAAGACGCGCTCATCCCGTGCAGG 480

      85 AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLe 101
         |||   :::::::|||        :::  :::   |||   :::|||||
     481 GCTCCAGTTTCCAGTGGGACGTGTTCATCGGTTTCTCCGAAAGGGCAACT 530

     101 uLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnL 118
         |          :::|||   |||:::::::|||   |||
     531 ATGCAAAGAGGGTCGGCGGTGGAGCTCCTGTCTACAT.........GGCT 571

     118 euProPro.GluHisLeuTrpGlyGlnIleValProThrLeu...HisLe 133
               ||| |||:::|||   :::::::|||:::            :::
     572 GCCGTCCTAGAGTACCTCACTGCCGAAATCTTGGAACTCGCAGGCAACGC 621

     133 uTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleA 150
           :::|||   |||:::      |||:::|||   :::   |||:::
     622 TGCCCGCGACAACAAGAAATCTAGGATCATCCCACGCCACCTTCAACTC. 670

     150 rgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAla 166
           :::|||   ::::::   |||||||||:::      ||||||::::::
     671 ..GCTGTGCGTAATGATGAAGAACTCAACAAGCTTTTGGGTGGGGTGACG 718

     167 MetSerLys 169
         :::::::::
     719 ATCGCTCAA 727

seq_name: em_htgo_inv:AC034780

seq_documentation_block:
ID   AC034780   standard; DNA; HTG; 1066 BP.
XX
AC   AC034780;
XX
SV   AC034780.1
XX
DT   07-APR-2000 (Rel. 63, Created)
DT   07-APR-2000 (Rel. 63, Last updated, Version 1)
XX
DE   Giardia intestinalis clone AJ1492 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-1066
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-1066
RA   McArthur A.G., Morrison H.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (05-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
```

```
CC   * be preserved.
CC   *        1     1066: contig of 1066 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .1066
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="AJ1492"
XX
SQ   Sequence 1066 BP; 312 A; 270 C; 224 G; 259 T; 1 other;


alignment_scores:
            Quality:   97.50              Length:    241
              Ratio:  0.871                Gaps:     12
 Percent Similarity:  46.473   Percent Identity: 22.822

alignment_block:
 US-09-674-779-2 x AC034780   ..

 Align seg 1/1  to: AC034780  from: 1  to: 1066

       4 PheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuValAlaCy 20
         :::|||::::::      :::         ::::::   :::|||   ||
     376 TATAATGAGTTTGTCAAGAGCAAGTACAGGGATAACCCCGTTGTTCTCTG 425

      20 s.........SerAlaProIleProThrAsnProGlnValSerProIleL 34
         |         |||:::   |||   :::   ||||||   |||      :
     426 TCAGTGTCTTTCTAGTGGTATTGACAGCCTTCCTCAGAGCTCCAAGCCGT 475

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         ::|||||||||||||   ||||||:::   :::   ||||||:::   :::
     476 CAACCCCTTCCGTACGGATTACATCTGGTAGT...GGAGATTATTCTGTA 522

      51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
            |||::::::      ::::::|||::::::
     523 ...GAGAGAAACATTATGCTAACACACCTGAAC..................  552

      67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGlu.............  79
                       |||:::   |||:::   :::
     553 ................CATAATGACACGAATGCTAGCCCTGGCTCCGCCC 586

      80 ........ValValArgTyrGlnAlaTyrLeuGlnSerArgLeu......  91
                :::|||||||   |||      :::|||||||:::|||
     587 TTTCTCTCATCGTGAGGACACAAAGAGCAATCCAGTCTAAGCTTAGTGCC 636

      92 .........GlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAl 105
                 ::::::      |||   ||||||:::   |||
     637 ATGAAAACATCGCATACCCAGCCAGATATGTCGTCTTCTCTG........ 678

     105 aArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluH 122
                            |||:::        :::         |
     679 .....................CACGATTCCCTAGGTATTAAGACTATTC 706

     122 isLeuTrp..................GlyGlnIleValProThrLeuHis 132
         ||   |||                 ::::::|||      |||||||||
     707 ATCCTTGGGTTCGGGGGCACTCTTCCTCTCGTATTTCAACAACCCTGCAC 756

     133 Leu.........TyrGlnAspLeuLysSerArgGlyIleLeuProAlaAs 146
         |||         :::   :::   ::::::   ||||||   |||   ||
     757 TTGGGAGATTGTCACATAGAAGAATCTGAGATAGGCATCGCTCCCCTCAA 806

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
         ||||:::   :::            |||   ::::::::::::||||||
     807 TACAGACAGGCAA......ATGTCAAACTCTCAAGTTGATGACTGTGCCG 850

     163 ly..............................GlyAlaAlaMetSer 168
                                        ::::::::::|||:::
     851 ATACCCAGGCCTTCCGTGAATTCACTCTCAGCACGTCCACTAGCATGGCC 900

     169 LysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIl 185

         |||   :::|||:::         :::         :::   :::
     901 TTGCATTCAGCTAATGAACACGRGCATATGTCTAAGGAGAGCTGCGACTA 950

     185 eLysSerGlnAla..............LeuTyrGluLeuGlnAsnArgL 197
           |||   |||                :::   ||||||:::|||
     951 TGACAGCCCGGCAGTCGCCCTTATCGTTATGAAGGAGCTAGAGAATCTAT 1000

     197 euCysGlnTyrTrpLeuGluHis 204
           :::      ::::::|||
    1001 CGAATAAAAGCCTGGTTGACCAT 1023
```

OM of: US-09-674-779-2 to: N_Geneseq_032802:* out_format : pfs

Date: Sep 16, 2002 6:57 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124122_25875/app_query.fasta_1.310
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=std.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 199.900000

score_list:

| Sequence | Strd | Orig | ZScore | EScore | Len | ! Documentation .. |
|---|---|---|---|---|---|---|
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59797 | + | 1355.00 | 2647.89 | 3.8e-1 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF28530 | + | 1355.00 | 2609.26 | 5.5e-1 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL20561 | - | 98.50 | 152.39 | 3.85 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL20560 | + | 98.50 | 139.23 | 20.85 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL11490 | + | 98.50 | 138.85 | 21.88 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL11502 | + | 98.50 | 138.60 | 22.59 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS36567 | + | 95.00 | 153.50 | 3.34 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI59893 | - | 94.00 | 158.76 | 1.70 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL05039 | - | 93.50 | 127.50 | 93.81 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS45146 | - | 92.00 | 150.56 | 4.87 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV33474 | + | 92.00 | 128.95 | 77.88 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH26027 | + | 89.50 | 147.11 | 7.58 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL15895 | + | 89.50 | 135.16 | 35.10 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL15779 | + | 89.50 | 133.61 | 42.85 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL15778 | - | 89.50 | 129.73 | 70.47 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS35758 | + | 89.50 | 118.46 | 299.15 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS35754 | + | 89.50 | 118.42 | 300.68 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS35759 | + | 89.50 | 112.94 | 607.25 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK79227 | + | 89.50 | 112.94 | 607.25 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAD10354 | - | 89.00 | 140.54 | 17.61 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL02149 | + | 88.50 | 146.30 | 8.41 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS30562 | + | 88.50 | 144.63 | 10.42 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL06616 | + | 88.50 | 144.63 | 10.42 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS80217 | + | 88.00 | 149.99 | 5.24 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH81475 | + | 88.00 | 149.22 | 5.78 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH84500 | + | 88.00 | 149.22 | 5.78 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS90352 | + | 88.00 | 149.06 | 5.90 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL03309 | + | 88.00 | 148.64 | 6.23 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS46271 | - | 88.00 | 120.46 | 231.35 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA82625 | - | 88.00 | 101.43 | 2.7e+0 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK94427 | - | 87.00 | 142.96 | 12.90 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK80624 | + | 87.00 | 126.91 | 101.15 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK80625 | + | 87.00 | 126.89 | 101.35 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL25734 | - | 86.50 | 130.61 | 62.91 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV18082 | - | 86.00 | 141.18 | 16.22 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:AAX77394 | - | 86.00 | 139.46 | 20.22 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:AAN81083 | - | 86.00 | 138.89 | 21.75 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS75095 | + | 86.00 | 137.06 | 27.52 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:ABL32230 | - | 86.00 | 114.67 | 485.81 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT11658 | - | 86.00 | 109.79 | 909.37 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:ABA05882 | - | 86.00 | 109.79 | 909. | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI99683 | - | 86.00 | 50.32 | 1.4 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI99682 | - | 86.00 | 50.30 | 1.4 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ53602 | + | 85.50 | 144.87 | 10.1 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:AAQ46887 | + | 85.50 | 135.05 | 35.6 | | |

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59797

seq_documentation_block:
ID AAF59797 standard; DNA; 753 BP.
XX
AC AAF59797;
XX
DT 04-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis strain ATCC43617 BASB120 DNA.
XX
KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory; ds.
XX
OS Moraxella catarrhalis.
XX
PN WO200109335-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07361.
XX
PR 03-AUG-1999; 99GB-0018281.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2001-159872/16.
DR P-PSDB; AAB60645.
XX
PT New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
PT strain American Type Culture Collection 43617, for use as therapeutic
PT agents or vaccines against bacterial infections, e.g. otitis media or
PT pneumonia -
XX
PS Claim 13; Page 64; 75pp; English.
XX
CC The invention relates to the Moraxella catarrhalis strain ATCC43617
CC BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
CC invention also relates to immunogenic fragments of the BASB120 protein,
CC expression vectors and host cells comprising BASB120 nucleic acids, the
CC recombinant production of BASB120, vaccine compositions comprising the
CC BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
CC compositions comprising the anti-BASB120 antibody, and a method of
CC identifying a Moraxella catarrhalis infection via the detection of
CC BASB120 proteins or antibodies. The vaccine compositions of the invention
CC are useful as prophylactic or therapeutic agents against Moraxella
CC catarrhalis infections in mammals, particularly humans. Moraxella
CC catarrhalis is a Gram negative bacterium frequently isolated from the
CC human upper respiratory tract, which is responsible for several
CC pathological conditions. It is responsible for about 15% of otitis media
CC cases in children (which can lead to temporary or permanent hearing
CC loss). It also causes pneumonia in elderly people, and sinusitis,
CC nosocomial infections and, less frequently, invasive diseases. BASB120
CC proteins or nucleotides may additionally be used in screening for novel
CC antibacterial compounds, and in the diagnosis and staging of infections.
CC The present sequence represents DNA encoding the Moraxella catarrhalis
CC strain ATCC43617 BASB120 protein.
XX
SQ Sequence 753 BP; 239 A; 166 C; 152 G; 196 T; 0 other;

alignment_scores:
Quality: 1355.00 Length: 250

```
            Ratio:  5.420                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF59797    ..

 Align seg 1/1  to: AAF59797  from: 1  to: 753

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 50

      17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 100

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 150

      51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 200

      67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 250

      84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 300

     101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 350

     117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 400

     134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 450

     151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 500

     167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 550

     184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 600

     201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 650

     217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 700

     234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 750

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF28530

seq_documentation_block:
ID   AAF28530 standard; DNA; 23210 BP.
XX
AC   AAF28530;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #17.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 141-146; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 23210 BP; 7108 A; 4482 C; 4951 G; 6668 T; 1 other;


alignment_scores:
            Quality: 1355.00              Length:    250
              Ratio:  5.420                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF28530    ..

 Align seg 1/1  to: AAF28530  from: 1  to: 23210

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11483 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 11532

      17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11533 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 11582

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11583 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 11632

      51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   11633 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 11682

      67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||
  11683 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 11732

     84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11733 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 11782

    101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11783 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 11832

    117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11833 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 11882

    134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11883 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 11932

    151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11933 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 11982

    167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11983 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 12032

    184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12033 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 12082

    201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12083 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 12132

    217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12133 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 12182

    234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12183 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 12232
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL20561

seq_documentation_block:

```
ID   ABL20561 standard; DNA; 4560 BP.
XX
AC   ABL20561;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 13156.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 13156; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4560 BP; 1344 A; 1186 C; 1089 G; 941 T; 0 other;
```

alignment_scores:

```
           Quality:   98.50              Length:    221
             Ratio:  0.895                Gaps:     12
 Percent Similarity:  49.774   Percent Identity: 25.792
```

alignment_block:

US-09-674-779-2 x ABL20561/rev ..

Align seg 1/1 to reverse of: ABL20561 from: 1 to: 4560

```
     17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
        |||||||||   :::::::|||         :::|||:::   ||||||
   2455 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 2411

     33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                                  :::|||:::   :::|||   |||
   2410 .....CGACGGGATCGTAATGCTGATGGGCATC 2383

     50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
          :::|||      :::   ::::::   :::   |||:::|||   :::
   2382 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 2339

     67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
        :::   |||:::      :::   :::   :::|||:::|||
   2338 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 2289

     81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         |||:::|||              |||      :::|||   ||||||
   2288 TAGGTGGCAG.............CGTCATCGAAGCTATCCTCCACCT. 2255

     98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                        |||:::::::|||:::
   2254 ................................GCAGCGGCGCACGAC 2240

    115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
        |||   ::::::||||||   |||   :::|||   :::   :::
   2239 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 2193

    129 ..........................ProThrLeuHisLeuTyrG 135
                                   |||:::            |
   2192 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 2143

    135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
        ||   :::::::::   :::   ||||||      :::|||   ||||||
   2142 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 2093

    152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
            :::   |||:::      :::   |||   |||   |||:::::
   2092 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 2043

    168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
```

```
      :::|||   |||:::|||:::    |||:::   |||         :::
   2042 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 1993

    185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
              |||   |||   :::::::|||      |||   |||:::
   1992 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 1949

    201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
              :::::::::::::|||   |||      |||   ||||||
   1948 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 1899

    215 yrAlaThrGly 218
          |||   |||
   1898 AGGCGATGGGC 1888
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL20560

seq_documentation_block:
ID ABL20560 standard; DNA; 14664 BP.
XX
AC ABL20560;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 13153.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 13153; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 14664 BP; 4286 A; 2912 C; 2978 G; 4488 T; 0 other;

alignment_scores:
Quality: 98.50 Length: 221
Ratio: 0.895 Gaps: 12
Percent Similarity: 49.774 Percent Identity: 25.792

alignment_block:

US-09-674-779-2 x ABL20560 ..

Align seq 1/1 to: ABL20560 from: 1 to: 14664

```
     17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
        |||||||||  :::::::|||         :::|||:::   ||||||
   4215 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 4259

     33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                                    :::|||:::   :::|||   |||
   4260 .....CGACGGGATCGTAATGCTGATGGGCATC 4287

     50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
        :::|||      :::   ::::::  :::   |||:::|||   :::
   4288 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 4331

     67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
        :::   |||:::      :::   :::   :::|||:::|||
   4332 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 4381

     81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         |||:::|||              |||      :::|||   ||||||
   4382 TAGGTGGCAG.............CGTCATCGAAGCTATCCTCCACCT. 4415

     98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                        |||::::::|||:::
   4416 ................................GCAGCGGCGCACGAC 4430

    115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
        |||   ::::::||||||   |||   :::|||   :::   :::
   4431 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 4477

    129 ..............................ProThrLeuHisLeuTyrG 135
                                       |||:::            |
   4478 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 4527

    135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
        ||   :::::::::   :::   ||||||      :::|||   ||||||
   4528 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 4577

    152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
            :::   |||:::`    :::   |||   |||   |||:::::
   4578 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 4627

    168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
        :::|||   |||:::|||:::    |||:::   |||         :::
   4628 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 4677

    185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
              |||   |||   :::::::|||      |||   |||:::
   4678 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 4721

    201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
              :::::::::::::|||   |||      |||   ||||||
   4722 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 4771

    215 yrAlaThrGly 218
          |||   |||
   4772 AGGCGATGGGC 4782
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL11490

seq_documentation_block:
ID ABL11490 standard; cDNA; 15166 BP.
XX
AC ABL11490;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 28952.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

```
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB67387.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 28952; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 15166 BP; 4466 A; 2981 C; 3067 G; 4652 T; 0 other;


alignment_scores:
            Quality:   98.50              Length:    221
              Ratio:  0.895                Gaps:     12
 Percent Similarity:  49.774   Percent Identity: 25.792

alignment_block:
 US-09-674-779-2 x ABL11490    ..

 Align seg 1/1  to: ABL11490  from: 1  to: 15166

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||||||||   :::::::|||          :::|||:::   ||||||
    4916 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 4960

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                               :::|||:::   :::|||   |||
    4961 .....................CGACGGGATCGTAATGCTGATGGGCATC 4988

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
         :::|||      :::   ::::::   :::   |||:::|||   :::
    4989 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 5032

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
         :::   |||:::      :::   :::   :::|||:::|||
    5033 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 5082

      81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         |||:::|||               |||      :::|||   ||||||
    5083 TAGGTGGCAG..............CGTCATCGAAGCTATCCTCCACCT. 5116

      98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                          |||:::::::|||:::
    5117 .............................GCAGCGGCGCACGAC 5131

     115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
         |||   ::::::||||||   |||   :::|||   :::   :::
    5132 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 5178

     129 ..........................ProThrLeuHisLeuTyrG 135
                                     |||:::          |
    5179 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 5228

     135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
         ||   :::::::::   :::   ||||||      :::|||   ||||||
    5229 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 5278

     152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
             :::   |||:::      :::   |||   |||   |||:::::
    5279 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 5328

     168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
         :::::|||   |||:::|||:::   :|||:::   |||         :::
    5329 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 5378

     185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
                 |||   |||   :::::::|||      |||   |||:::
    5379 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 5422

     201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                 :::::::::::::|||   |||      |||   ||||||
    5423 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 5472

     215 yrAlaThrGly 218
         |||   |||
    5473 AGGCGATGGGC 5483

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL11502

seq_documentation_block:
ID   ABL11502 standard; cDNA; 15504 BP.
XX
AC   ABL11502;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 28988.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB67399.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 28988; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
```

```
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 15504 BP; 4552 A; 3065 C; 3141 G; 4746 T; 0 other;


alignment_scores:
            Quality:   98.50              Length:    221
              Ratio:  0.895                Gaps:     12
 Percent Similarity:  49.774   Percent Identity: 25.792

alignment_block:
 US-09-674-779-2 x ABL11502   ..

 Align seq 1/1  to: ABL11502  from: 1  to: 15504

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||||||||   ::::::|||         :::|||:::   ||||||
    4916 CTGGTCGCATCGGAGTCTCCG...GCTTGCGGGCCTGAGAGTTCACCT.. 4960

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
                                 :::|||:::   :::|||   |||
    4961 .....CGACGGGATCGTAATGCTGATGGGCATC 4988

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
         :::|||      :::   ::::::  :::   |||:::|||   :::
    4989 TTCGGGAG......CGCATTGCGACATGGGCGGAGTTGGAGGCATTATAT 5032

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal......Va 81
         :::   |||:::      :::   :::   :::|||:::|||
    5033 CATCCATGGGTATGAGGACGGGATTCTTGAAGGCAATCTGTAAGCTTGTG 5082

      81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
          |||:::|||               |||      :::|||   ||||||
    5083 TAGGTGGCAG.............CGTCATCGAAGCTATCCTCCACCT. 5116

      98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                                        |||:::::::|||:::
    5117 .................................GCAGCGGCGCACGAC 5131

     115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
         |||   ::::::||||||   |||   :::|||   :::   :::
    5132 CCT...GAAGTCCCGCCAACTCATTTTGTATGGTATTCTTGTACGTATCC 5178

     129 ............................ProThrLeuHisLeuTyrG 135
                                     |||:::           |
    5179 CTCTGAACTTTGAAGTCCGTGTATATCTCATCCAGTTCGGCCACGGTCTC 5228

     135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
         ||   :::::::::   :::   ||||||      :::|||   ||||||
    5229 AAGGAGTTCGTCAAGGGTCTTATCTGCCTTGGATATCTCAAGGGCGGAGT 5278

     152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
             :::   |||:::      :::   |||   |||   |||:::::
    5279 CCAAAACATCTGCCTCAGCAGTGGGACTGGGCTCATGGACTTGCTGTTCA 5328

     168 rLysHisLeuThrAsnSerAlaIle.AspIleTrpValProAspLeuGlu 184
         ::::|||   |||:::|||:::    |||:::   |||          :::
    5329 ATCTCACTGGACTTCATCCACTCGGTGATCTTATCGTTGGCCGTCAAAGG 5378

     185 IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr.. 200
                |||   |||   :::::::|||       |||   |||:::
    5379 ACGTGGCTCCAACTGCTGACCCAAATCCAG......TTGCTCCAGTTCTC 5422

     201 .......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
             :::::::::::::|||   |||      |||   ||||||

    5423 GTACCTTTTCATCGATAATGGTGTCAATCTCTTTTTCCTGCTCGGATCGC 5472

     215 yrAlaThrGly 218
         |||   |||
    5473 AGGCGATGGGC 5483

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS36567

seq_documentation_block:
ID   AAS36567 standard; DNA; 2240 BP.
XX
AC   AAS36567;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human cardiovascular system antigen genomic DNA SEQ ID No 2067.
XX
KW   Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; tissue regeneration;
KW   anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155321-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01340.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
```

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451930/48.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
PS Claim 1; SEQ ID No 2067; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

alignment_scores:
Quality: 95.00 Length: 303
Ratio: 0.772 Gaps: 13
Percent Similarity: 40.594 Percent Identity: 19.802

alignment_block:
US-09-674-779-2 x AAS36567 ..

Align seg 1/1 to: AAS36567 from: 1 to: 2240

```
 17 LeuValAlaCysSerAlaProIlePro..................... 25
    |||   |||||||||:::      |||
992 CTTGAAGCCTGTTCTTCCGAAGAGCCTCACAGTGTGTAGCTCTTGGGCCT 1041

 26 ...............ThrAsnProGlnValSerProIleLysThrProS 37
                   :::   |||   :::|||::: .     ||||
1042 GTGTCAGGCCCTAAGGCTCTCCTGCCAGGAAGGACCACTCATCCCTCCAT 1091

 37 erValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHis 53
    |||||       :::|||   |||   :::::::|||   |||:::|||
1092 CTGTTCCCCAACCCCAGGAC...ATCCGCAACCAGCACAGACACAGGCAC 1138

 54 AspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLe 70
      ::::::::::::   |||   |||||||||   ::::::
1139 AGGCGGAAGGCAGAGCTGGTGAAGCTGCAGGCCACATTACAGGGCCTGAG 1188

 70 uGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrL 87
                    :::::::::         |||:::   |||:
1189 CACTAAGACCACCTTCTATGAGGAGCAGGGTGACTACTACAGCCAGTACA 1238

 87 euGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThr 103
    ::::::::   |||   ::::::|||   |||   |||:::   :::
1239 TCCGGGCCTGCCTG...GACCACCTGGCCCCCGACTCCAAGTAGGTGCTG 1285

104 ThrAlaArgSerTrpGln...AlaCysGlyHisGluProTyrGlnLeuPr 119
      :::|||:::|||:::   :::|||       :::|||       :::
1286 CTTTCCCGTACTTGGAAGGTTACGTGTGACTTCTCACCCCTGCCCATGTC 1335

119 oPro..............GluHisLeuTrpGlyGlnIleValProThrL 131
     |||                |||   |||:::      |||:::
1336 ACCTTTGCCCTGCTATCTCCCAGCTTTGAAAGGGGAATCTCACCCAGCAC 1385

131 euHis.....................LeuTyrGlnAspLeuLysSerArg 140
      |||                    :::         |||:::|||
1386 AGCACAGTCCCTTCCTAAGTCTGGTCTTTGAGCCTCATTTACAAAGTCTT 1435

141 GlyIleLeu............ProAlaAsnThr.GlnIleArgSerValT 153
    |||:::              |||   :::::: ::::::||||||:::
1436 GGGCTCAAGTGCAGCCCCCCTCCTCTGTCTTCCATCAGTGAGAAGTCTTC 1485

153 yrArg........................................... 154
      |||
1486 CTCGGGGTCTCATTTCCTTTCCAGCCACAGAAAACAAGTTTCTTTGGTGG 1535

155 ........................AsnProGluLeuAsnGln..... 160
                           :::|||::::::||||||
1536 TAAAGTGTTTATTGGATGAATTCTCTGGGGCCTTCTTTTAATCAGGAGTT 1585

160 ........................................... 160

1586 CTGGGAAGGGGAAGAAGCAGCCTTCTCTTCATTACACTGCTGCTCAGCTC 1635

161 ..........CysAlaGlyGlyAlaAlaMetSerLysHisLeu...... 171
               |||   ||||||         |||:::::|||
1636 CTGGAAAAGGGTGTCTTGGTGGAAATTGAAGATCTTCCCGCCTCTCAGTA 1685

172 .....................................ThrAsnSerAl 175
                                        :::|||   ::
1686 TGTGTCTTTGGAGGGCAGGATGTCAGCCCCTTCCTGTCCTCTAACCTGGG 1735

175 aIleAspIleTrpVal.............................ProA 182
    ::::::::::|||:::                            ||||
1736 GGTGGAGCTCTGGGCTGGGCACCGGGTGGAAGAGAAGAAGCATGGTCCCG 1785

182 spLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCys 198
    ||||||||      ::::::   :::   |||
1786 ATCTTGAAGGAGCTGCTAGAATGATGGTGGAG.................. 1817

199 GlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTy 215
              ||||||:::::::::::::::   ||||||      ::
1818 .........ACAGAGCATGCCCAGACAGAAAGTGAAGGGCTGTGGACATG 1858
```

```
    215 rAlaThr 217
        :::::::
   1859 GTCTAGC 1865

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI59893

seq_documentation_block:
ID   AAI59893 standard; cDNA; 1179 BP.
XX
AC   AAI59893;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 3882.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM40737.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 3882; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1179 BP; 218 A; 330 C; 328 G; 299 T; 4 other;
```

```
alignment_scores:
            Quality:   94.00              Length:    200
              Ratio:  1.000                Gaps:     12
 Percent Similarity:  47.000   Percent Identity: 27.500

alignment_block:
 US-09-674-779-2 x AAI59893/rev   ..

 Align seg 1/1  to reverse of: AAI59893  from: 1  to: 1179

      75 AlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGln..... 88
         ||||||   |||:::      :::   |||:::   ||||||
    1131 GCAACCACACAATCACCCTACAAGGTGCAAGGACAACTCCAGGGAGTCTC 1082

      89 ..................SerArgLeuGlyAsnTyrLeuProProMetS 99
                           |||::::::      |||   ||||||
    1081 AAGTGCAGCCTGGTGGCTCAGCCAGATTTTGGCATACACCCCACCCTACC 1032

      99 erGlnLeu.............................LeuThrThr 104
         :::                                     |||
    1031 CTGAGGCACCAGAGGGGGAAGGAACGGGACTAGGGGGCAGAGAGCTAACC 982

     105 AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGl 121
         :::|||:::   |||:::   |||      |||   :::      |||
     981 AGCAGAAAGGGGCAGGGACCGGGC...CTTCCCAACAAAACCTCCCCAAC 935

     121 uHisLeuTrpGly.GlnIle.........ValProThrLeuHisLeuTyr 134
                 ||| |||:::            ||||||   ||||||
     934 CTGGTGCGTTGGTGCAGCTGAGGCTCAGGAAGCCCACTCACCACTTGGTA 885

     135 GlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSe 151
         |||   |||||||||      :::      |||   :::|||      ::
     884 CAGCCGCTCAAATCAATGCCAGTCTGG...GCCTGCGCACAGCGGATGGC 838

     151 rValTyrArgAsnProGluLeuAsnGlnCys......AlaGlyGlyAlaA 166
         :|||                :::   :::   |||      |||   |||:::
     837 GGTA............CGCACCAGCACCTGCGGGTCAGCCTGGGGGTCGA 800

     166 laMetSerLysHis..............LeuThrAsnSerAlaIleAsp 177
              |||::::::                |||::::::|||   :::
     799 GGCCATCCAGGGAAGGAGACCATTCACCCCCAACCAGCACTGCCTCCTCT 750

     178 IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGl 194
         :::::::|||:::   :::   |||:::::::::         :::||
     749 TCTTTCCTGCCCAGCAAAAACTGCAAGAAGAGATCA......GGGATACA 706

     194 nAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheG 211
         ||||:::      :::|||                          :::|
     705 AAACAAAAAACTACATTAC.......................CTGG 683

     211 lyLeuGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe 227
         |||||   |||   :::   :::|||   |||   |||   :::|||
     682 GTCTCCCACTGCAGGGCAAGGCTGCCACCCATCAGGAT...GAAGGCAGC 636

     228 ArgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
         ||||||||||||::         :::      |||||||||:::|||
     635 CGGAAGTGGGGAAGTCCACCACACAGTGAGGTCTCTATCTGCAAACAT 588

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL05039

seq_documentation_block:
ID   AAL05039 standard; DNA; 17324 BP.
XX
AC   AAL05039;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 7727.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
```

```
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition -
XX
PS   Disclosure; SEQ ID NO 7727; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 17324 BP; 3826 A; 5147 C; 4719 G; 3632 T; 0 other;


alignment_scores:
            Quality:   93.50              Length:    226
              Ratio:  0.926                Gaps:     11
 Percent Similarity:  44.690   Percent Identity: 23.451

alignment_block:
 US-09-674-779-2 x AAL05039/rev   ..

 Align seg 1/1  to reverse of: AAL05039  from: 1  to: 17324

      58 SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHi 74
         :::|||   |||:::|||         |||:::   |||   ::::::
    7074 ACCCACCAGGGGATCCAGTTCTTGGTGGAGAATGAACTGCTGCAGAAC.. 7027

      74 sAlaThrLysGlnGluValValArgTyr..................... 83
           |||   :::|||::::::|||:::
    7026 ....ACACCCGAGGAGATCGCCCGCTTCCTGTACAAGGGCGAGGGGCTGA 6981

      84 .............GlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr... 94
                        ||||||   :::|||   |||::::::
    6980 ACAAGACAGCCATCGGGGACTACCTGGGGGAGAGGTACGGTCACCACTCA 6931

      95 .............LeuProProMetSerGlnLeuLeuThrThrAlaAr 106
                      |||||||||      |||
    6930 GCTCCTGTGGGGCCCCTCCCTCCCACCCCCCAGACCCAGCTCCTGCCCTC 6881

     106 gSerTrpGlnAlaCysGly.HisGluProTyrGlnLeuProProGluHis 122
           ||||||      ||| ::::::|||   :::||||||   |||
    6880 ACACTGGCAGCTCACAGGTGGAACAGCCCTGCAGCCTCCCCGCAGAGGGG 6831

     123 LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSe 139
           |||                  :::   |||         |||   ::
    6830 CCCTGG.............CAGATCAGCCTTCTGACTGGACTGTGACA 6796

     139 rArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnP 156
         :::|||:::            |||      |||::::::      :::|
    6795 GAATGGACTG...........ACATGGGCTAGGGGAGCCTGCGGCCAAC 6758

     156 roGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
         ||:::   :::     :::|||  ::::::::::::   :::|||
    6757 CGAGAGCATCTGGGGATGTGGGGCCTTCTGGAATTACCTGGCAGTTGTGG 6708

     173 AsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAl 189
         :::|||   |||   :::   :::      :::|||      ||||||::
    6707 GGAAGC...ATCCATTTATGTCTGAGAAGGGTAGAGAACATCTCTCAAGG 6661

     189 aLeuTyrGluLeuGlnAsnArg.............LeuCysGlnTyrT 201
         :   :::|||||||:::|||||||          :::   :::::::|
    6660 ATCATGGGAGCTTGAGAACCGTGGGTAGACCATGGCTGTGCATAGGCACT 6611

     201 rpLeuGluHisGlyGlu...................................... 206
         ||:::        |||
    6610 GGGTGGCCATCTGGGAGCTGTGGGGACCGCTGGGCCCTTGTGGGAGCTTA 6561

     207 .....................................AsnGl 208
```

```
                                                          :::||
    6560 GGAGTTGTGGGGGTCATTTGGAAGCCTAGGGTGCACAGGACTGGACATCA 6511

     208 nAsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrG 225
         |   :::|||      |||::::::::::|||               :::|
    6510 GATGATTGGAGGCCATCTGTGGACGTCTGGG.............AATC 6476

     225 lnGlyPheArgLys...TrpGlyAla 232
         ||||||||   :::   |||||||||
    6475 AGGGATTTATAGAGATCTGGGGTGCT 6450

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS45146

seq_documentation_block:
ID   AAS45146 standard; cDNA; 1721 BP.
XX
AC   AAS45146;
XX
DT   18-DEC-2001  (first entry)
XX
DE   cDNA encoding novel human secretory protein, Seq ID No 415.
XX
KW   Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW   ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW   transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW   amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW   ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW   gut protection; lung; liver fibrosis; immune deficiency; infection;
KW   severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW   multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW   fertility; analgesic; pain; antigen; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166689-A2.
XX
PD   13-SEP-2001.
XX
PF   05-MAR-2001; 2001WO-US04942.
XX
PR   07-MAR-2000; 2000US-0519705.
PR   19-MAY-2000; 2000US-0574454.
PR   17-JUN-2000; 2000US-0596193.
PR   14-JUL-2000; 2000US-0616847.
PR   19-SEP-2000; 2000US-0665363.
PR   20-OCT-2000; 2000US-0693267.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Xu C,  Wehrman T,  Ren F,  Ma Y,  Zhou P;
PI   Zhao QA,  Yang Y,  Drmanac RT,  Zhang J,  Chen R,  Xue AJ,  Wang J;
XX
DR   WPI; 2001-589934/66.
DR   P-PSDB; AAU28246.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis and treatment of
PT   cancer, neurological, inflammatory, and autoimmune disorders -
XX
PS   Claim 1; SEQ ID No 415; 107pp; English.
XX
CC   The invention relates to novel isolated human secreted polypeptides (I)
CC   and polynucleotides (II). (I) and (II) are useful for treating
CC   inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC   ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC   involved in increasing haematopoiesis, stem cell survival, bone growth
CC   and remodeling. (I), (II) and modulators of (II) are useful for
CC   prophylaxis or treatment of one or more cancers. (II) is also useful for
CC   creating transgenic animals useful for studying the in vivo activities of
CC   the polypeptide as well as for studying modulators of the polypeptides.
CC   (I) induces the proliferation of neural cells and regeneration of nerve
CC   and brain tissue and is useful for the treatment of central and
CC   peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC   Parkinson's disease, Huntington's disease, and amyotrophic lateral
CC   sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis and is useful for treating myeloid
CC   or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC   and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC   tissue growth, and in tissue repair, healing of burns, incisions,
CC   ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
CC   disorders, or periodontal disease. Furthermore, (I) is also useful for
CC   gut protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, various immune deficiencies and
CC   disorders including severe combined immunodeficiency (SCID), bacterial or
CC   fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC   rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC   reactions and conditions, such as asthma or other respiratory problems.
CC   In addition, (I) affects biorhythms or circadian cycles of rhythms,
CC   fertility, metabolism, catabolism, anabolism, storage or elimination of
CC   dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC   analgesic effects or other pain reducing effects, immunoglobulin like
CC   activity and can act as an antigen in a vaccine composition to raise an
CC   immune response. AAS44920-AAS45295 represent novel human secreted protein
CC   coding sequences of the invention.
XX
SQ   Sequence 1721 BP; 479 A; 373 C; 487 G; 381 T; 1 other;


alignment_scores:
            Quality:   92.00              Length:    152
              Ratio:  1.195                Gaps:     10
 Percent Similarity:  50.658   Percent Identity: 26.316

alignment_block:
 US-09-674-779-2 x AAS45146/rev   ..

 Align seg 1/1  to reverse of: AAS45146  from: 1  to: 1721

       5 AsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuValAlaCys.. 20
         |||::::::   |||   :::      |||::::::|||     |||
     837 AACAGATTCTCCATCATCTCCTGCTCCTCAGCCGTGCTGGGATTGTGTCT 788

      21 .......................SerAlaProIleProThrAsnProG 29
                                 |||:::   ||||||:::::
     787 TTTAAACACGGATAACTGCTGAAGAAGCACATCGATTCCATCCAGC.... 742

      29 lnValSerProIleLysThrProSerValLeuIleThrLysAspLysIle 45
           ||||||   :::  |||      :::|||:::   :::::
     741 .....TCCCCCAAGCAATTCCCTGTTTTCATCATTGTCCTGGAGCAATAT 697

      46 GlyAspHisHisThrHisGluHisAspGluSerValSerHisValGlyLe 62
         |||:::|||   |||              :::::::::   ||||||::
     696 GGCCAGCACTTCACT..............GCAATACAGTTTGTTGGCAT 662

      62 uGlnAlaHisPheGluThrTrpLeuGln...MetHisHisAlaThrLysG 78
         ::::   |||:::        ||||||           ||||||   :
     661 CAAAAGGCATCTT......TGCCTTCAGCCCTCTTCAACAGCCACTGTAG 618

      78 lnGluValValArgTyrGlnAlaTyrLeuGlnSerArg..... 90
         :::::   :::   :::      |||   ::::::
     617 AAGACCCTGCTGGGCACCCTCTGTACACATCTCAGGCCGGAACTCAGCCA 568

      91 .......................LeuGlyAsnTyrLeuProProMetSe 99
                                 |||::::::||||||||||:::||
     567 TGTTTTCCACAATAGCCAGAGTGTTGTGGACGCCATCTGCCTCCTCTTTC 518

      99 rGln.LeuLeuThrThrAlaArg...SerTrpGlnAlaCysGlyHisGlu 114
         |||| |||:::   :::   :::      :::|||   |||   |||
     517 ACAGACTCATCCAGGCGCTCCAGATTCTGTACCAGCAGTGCTACCACCCT 468

     115 ProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLe 131
           :::      ||||||||||||      :::   :::::::|||   ||
     467 GCCCAT......CCACCAGAGCATCGATGAGCACTTCTGCTCCCT...CT 427

     131 uHis 132
         ||||
```

```
    426 TCAC 423

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV33474

seq_documentation_block:
ID   AAV33474 standard; cDNA; 11717 BP.
XX
AC   AAV33474;
XX
DT   10-DEC-1998  (first entry)
XX
DE   Girdwood S.A.virus cDNA sequence.
XX
KW   Bone marrow cell; alphavirus; Girdwood S.A.virus; cytokine; capsid;
KW   reverse transcriptase-PCR; nsP1; nsP2; nsP3; nsP4; E3; E2; 6K; E1;
KW   growth hormone; growth factor; interleukin; chemokine; enzyme;
KW   ribozyme; antisense oligonucleotide; ss.
XX
OS   Girdwood S.A. virus.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..59
FT                   /*tag=  a
FT   CDS             60..7616
FT                   /*tag=  b
FT                   /product= "Non-structural polyprotein"
FT                   /note= "Internal stop codon present at nucleotides
FT                          5763-5765"
FT   CDS             7662..11399
FT                   /*tag=  c
FT                   /product= "Structural polyprotein"
FT   3'UTR           11400..11717
FT                   /*tag=  d
XX
PN   WO9836779-A2.
XX
PD   27-AUG-1998.
XX
PF   18-FEB-1998;   98WO-US02945.
XX
PR   19-FEB-1997;   97US-0801263.
XX
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Davis NL,  Johnston RE,  Simpson DA;
XX
DR   WPI; 1998-495361/42.
DR   P-PSDB; AAW70462, AAW70463, AAW70473, AAW70474, AAW70475, AAW70476,
DR   AAW70477; P-PSDB; AAW70478, AAW70479.
XX
PT   Expression of heterologous RNA in bone marrow cells - using a
PT   recombinant alphavirus comprising a promoter operable in bone marrow
PT   cells operably associated with heterologous RNA.
XX
PS   Example 2; Fig 3A-3C; 68pp; English.
XX
CC   The invention provides a method for introducing and expressing
CC   heterologous RNA in bone marrow cells using alphavirus vectors.  The
CC   Girdwood S.A.virus is an alphavirus whose cDNA sequence was determined
CC   from uncloned reverse transcriptase-PCR reaction fragments amplified
CC   from the virion RNA.  The Girdwood S.A.virus cDNA encodes nonstructural
CC   and structural polyproteins.  The nonstructural polyprotein is,
CC   presumably, post-translationally modified into four different protein
CC   products, namely nsP1 (AAW70462), nsP2 (AAW70463), nsP3 (AAW70473), nsP4
CC   (AAW70474) proteins.  The structural polyprotein is, presumably, post-
CC   translationally modified into five different protein products, namely
CC   capsid (AAW70475), E3 (AAW70476), E2 (AAW70477), 6K (AAW70478), E1
CC   (AAW70479) proteins.  The Girdwood S.A.virus cDNA was used in the method
CC   of the invention.  The inventors claim the transformed bone marrow cells
CC   are useful for expressing a protein or peptide suitable for protecting
CC   the subject against a disease such as a microbial, bacterial, protozoal,
CC   parasitic or viral disease.  The transformed bone marrow cells are also
CC   claimed to be useful for expressing proteins and peptides such as
CC   hormones, growth hormones, growth factors, interleukins, cytokines,
CC   chemokines, enzymes, ribozymes or antisense oligonucleotides.
XX
SQ   Sequence 11717 BP; 3297 A; 3130 C; 2927 G; 2360 T; 3 other;


alignment_scores:
            Quality:   92.00              Length:    283
              Ratio:  0.724                Gaps:     18
 Percent Similarity:  44.876   Percent Identity: 24.735

alignment_block:
 US-09-674-779-2 x AAV33474    ..

 Align seg 1/1  to: AAV33474  from: 1  to: 11717

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||      |||   :::||||||           |||::
    8664 ACCTTGACCAGCCCGTACTTGGGCACATGCTCGTACTGTCACCATACTGA 8713

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
         :|||     ||||||||||||      :::|||         .   |||:
    8714 ACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGG.........GATG 8754

      44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
         ::      |||  ||||||||||||      :::   |||
    8755 AAGCGGACGACAACACCATACGCATACAGACTTCCGCCCAGTT....... 8797

      60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                  |||::::::      :::::::::
    8798 .........................TGGATACGACCAAAGCGGAGCAGC 8821

      77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
         |||   :::   ||| ||||||::::::      |||
    8822 AAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACCGTCA 8871

      88 ..................................................... 88

    8872 AAGAAGGCACTATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGA 8921

      89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
                |||      |||   :::|||           ||||||
    8922 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGA 8971

      98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
             ::::::      ||||||||||||||||      :::|||
    8972 CAGCGTAACGGTTAGTATAGCGAGTAGCAACTCAGCAACGTCATGCACAA 9021

     112 ............................GlyHisGluProTyrGlnLeu 118
                                     |||:::|||   |||:::|||
    9022 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGACCTA 9071

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
         ||||||   |||      |||:::   :::|||          :::|||::
    9072 CCTCCCGTTCAC......GGTAAGAAGATTCCTTGCACA...GTGTACGA 9112

     135 nAspLeuLys......SerArgGlyIleLeu............ProAla. 145
         :   ||||||      :::   |||   :::            |||:::
    9113 CCGTCTGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGAC 9162

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
         :::|||   ||||||   ::::::||||||   :::
    9163 CGCACGCCTATACGTCCTATCTGGAGGAATCATCAG.............. 9198

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
          |||   ::::::   |||   |||
    9199 .GGAAAGTCTACGCGAAGCCACCAT....................... 9222

     179 rpValProAspLeuGluIleLysSerGlnAla.LeuTyrGluLeuGlnAs 195
             |||:::      ::::::::::::||| :::   :::|||||||::
    9223 .....CCGGAAAGAACATTACGTACGAGTGCAAGTGCGGCGATTACAAGA 9267

     195 nArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212
```

```
      :|||         |||      :::      :::|||:::
 9268 CCGG.........TACCGTTACGACCCGTACCGAAATCAC.......... 9298

  212 euGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg 228
        |||||||:::         |||:::::::         :::      :::
 9299 ..GGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAAC 9346

  229 LysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArg 242
      :::   |||   |||||||:::      ::::::   ||||||
 9347 GAAGTGGGTCTTCAATTCGCCGGACTTGATCAGACATGCCGA 9388
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH26027

seq_documentation_block:

ID AAH26027 standard; DNA; 1510 BP.
XX
AC AAH26027;
XX
DT 05-SEP-2001 (first entry)
XX
DE DNA encoding Pseudomonas fluorescens pyoluteorin halogenase PltM.
XX
KW Pyoluteorin halogenase; regiospecific; halogenation; PltM;
KW pyrrolnitrin; transgenic plant; crop protection;
KW biological control; ds.
XX
OS Pseudomonas fluorescens.
XX
PN WO200144447-A1.
XX
PD 21-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-EP12347.
XX
PR 15-DEC-1999; 99US-0228801.
PR 03-JAN-2000; 2000US-0219343.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Steffens J, Batie C, Dietz JM, Dong J, Kamdar KP, Hill S;
XX
DR WPI; 2001-417902/44.
DR P-PSDB; AAB82510.
XX
PT Transferring a halogen to a substrate in a regiospecific manner for the
PT production of pyrrolnitrin for plant protection comprises contacting
PT the substrate with a regiospecific halogenase and an electron
PT transferase -
XX
PS Claim 12; Page 79-84; 194pp; English.
XX
CC The present sequence is that of DNA encoding PltM (see AAB82510), a
CC regiospecific pyoluteorin halogenase of Pseudomonas fluorescens.
CC The enzyme can be used for the biosynthesis of halogenated natural
CC products, where the halogenation is substrate and regiospecific. A
CC claimed method for transferring a halogen to a substrate in a
CC regiospecific manner comprises contacting the substrate with a
CC regiospecific halogenase (see AAB82506-13) in the presence of an
CC oxidant, a halogen donor, an electron transferase (see AAB82514-20),
CC and a reductant. A host cell that expresses at least 1 endogenous
CC regiospecific halogenase and which also includes a heterologous
CC nucleic acid encoding an electron transferase may be used.
CC Halogenated metabolites produced by these methods are used for the
CC protection of host organisms against pathogens. For example, crop
CC plants can be protected against fungal disease by expression of
CC pyrrolnitrin from a host cell. Transgenic plants and transgenic
CC microorganisms for the biosynthesis of halogenated natural products
CC are provided. Plants with enhanced resistance to phytopathogens
CC and biological control organisms with enhanced activity can be
CC obtained.
XX
SQ Sequence 1510 BP; 304 A; 503 C; 410 G; 293 T; 0 other;

```
alignment_scores:
            Quality:   89.50              Length:    304
              Ratio:  0.722                Gaps:     17
 Percent Similarity:  40.789   Percent Identity: 22.039

alignment_block:
 US-09-674-779-2 x AAH26027   ..

 Align seg 1/1  to: AAH26027  from: 1  to: 1510

   20 CysSerAlaProIleProThrAsnProGlnValSerProIleLysThrPr 36
      |||   :::|||   |||:::::::|||:::   |||            ||
  473 TGTCCAACGCCGCCCCCGTCAAGGCCGCGTTCATCATTGACGCTGCTGCC 522

   36 oSerValLeu.....IleThrLysAspLys.Ile 45
      |   :::|||                 :::   |||
  523 CAGGGCTCTCCGCTTTCCCGCCAACTGGGCTTGCGCACCACCGAAGGGCT 572

   46 GlyAspHisHis.....ThrHisGluHisAspGlu..... 55
      ||||||:::|||      |||   |||   :::
  573 GGCGACCGACACCTGCTCATTCTTCACCCACATGCTCAATGTGAAGAGCT 622

   56 .....SerValSerHisValGlyL 62
                 |||      |||      :
  623 ACGAAGATGCCCTGGCTCCGTTGTCCCGCACTCGTTCCCCCATCGAACTG 672

   62 euGlnAlaHisPheGluThrTrpLeuGln.....MetHisHis 74
      ::|||   |||:::         |||:::          :::  |||
  673 TTCAAGAGCACCTTGCACCACATCTTCGAAGAGGGCTGGTTGTGGGTCAT 722

   75 AlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLe 91
         :::|||      :::   :::|||:::   ::::||   |||||
  723 CCCCTTCAACAACCACCCGCAGGGCACCAATCAGTTGTGCAGCATCGGCT 772

   91 uGlyAsnTyrLeuProProMetSerGln..... 100
      |            |||::::::|||
  773 T.....CCAGTTCAACAACGCCAAGTACCGTCCCACCGAG 807

  101 .....LeuLeuThrThrAla..... 105
                :::     ||||||
  808 GCGCCGGAGATCGAGTTTCGCAAACTGCTGAAAAAGTACCCGGCCATCGG 857

  106 ArgSerTrpGlnAlaCysGlyHisGluPro..... 115
      |||:::   |||:::||||||:::   |||
  858 CGAACACTTCAAGGATGCGGTCAATGCCCGGGAGTGGATCTACGCGCCGC 907

  116 .TyrGlnLeuProProGlu...HisLeuTrpGlyGlnIleValProThrL 131
       :::|||||||||   :::          |||:::::::::|||:::
  908 GCATCAACTACCGCAGCGTGCAAAATGTCGGGGATCGCTTCTGCCTGCTG 957

  131 eu.....HisLeuTyrGln.AspLeuLysSerArgGlyIleLeuPr 144
            :::::::|||:::    |||   ||||||||||:::::
  958 CCGCAAGCCACAGGGTTTATCGACCCGCTGTTCTCCAGGGGGTTGATC.. 1005

  144 oAlaAsnThrGlnIleArgSerValTyrArg..... 154
         |||   ::::::|||:::   |||
 1006 .....ACCACCTTCGAGTCCATCCTCAGGCTGGCCCCCAAGGTGCTGG 1048

  154 ..... 154
 1049 ACGCCGCCCGCAGCAACCGCTGGCAACGGGAACAGTTCATCGAAGTCGAG 1098

  155 .....AsnProGluLeuAsnGlnCysAl 162
                  |||   ::|||   :::|||::
 1099 CGCCATTGCCTGAACGCGGTGGCGACCAATGACCAGTTGGTCTCCTGCTC 1148

  162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
      :      |||      ::   ||||||   |||           :::|
 1149 CTATGAAGCCTTCAGCGACTTTCACCTGTGGAACGTGTGGCATCGGGTCT 1198

  179 rpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsn 195
      ||:::         ::::::   |||      |||   :::|||   :::
```

```
   1199 GGCTCAGCGGCTCCAACCTGGGCAGTGCCTTTCTGCAAAAGCTGCTGCAC 1248

    196 ArgLeuCysGlnTyrTrpLeuGluHis...GlyGluAsnGlnAsnPheGl 211
                                 |||||||||   |||:::   ::::::|||
   1249 GAC..............CTGGAACACAGTGGCGACGCCCGCCAGTTCGA 1283

    211 yLeuGlyLeuTyrAlaThr.........GlyAlaIleHisLeuAspThrG 225
           :::|||   |||:::          |||::::::   |||||||:::
   1284 TGCAGCGCTTGAGGCGGTGCGCTTCCCTGGCTGCCTGTCCCTGGACTCGC 1333

    225 lnGlyPhe 227
          ::::::
   1334 CCGCCTAC 1341

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL15895

seq_documentation_block:
ID   ABL15895 standard; cDNA; 4360 BP.
XX
AC   ABL15895;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 42167.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB71792.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 42167; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4360 BP; 1134 A; 1166 C; 1129 G; 931 T; 0 other;


alignment_scores:
            Quality:   89.50              Length:    168
              Ratio:  1.065                Gaps:      8
 Percent Similarity:  50.000   Percent Identity: 24.405

alignment_block:

US-09-674-779-2 x ABL15895   ..

Align seq 1/1  to: ABL15895  from: 1  to: 4360

     43 AspLysIleGlyAspHisHisThrHisGluHisAspGlu..... 55
        |||:::::::|||   |||:::::::|||      ::::::
   1634 GACCGTCTTGGTGCTCACCGATCGCATTTACCAGAATCTCAACGGCCACA 1683

     56 ....SerValSerHisValGlyLeuGlnAlaHis..... 65
           ||||||      :::|||   :::::::|||
   1684 TGGATCAGTCTGCTCTGCTGGATCTGATGTACATGGCCAAGGTGAGCGAG 1733

     66 ..PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValVal 81
          |||         |||:::      |||:::      ::::::   :::
   1734 CCTTTCCTCATCGACTTGATAGCCCACACTCACCAGGATCGCGAAGCATT 1783

     82 ArgTyrGlnAlaTyr.................LeuGlnSerArgLeuGl 92
        |||::::::                       |||      :::   ::
   1784 CGACACAATTTTCAGTCAGGTGCTTCGTGGCCTCTTTGCCGGAATGCAGC 1833

     92 yAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpG 109
        :::::|||||      ::::::|||      ::::::      :::   |
   1834 GCAATATCTGCACTTCCAAGATCAGTGTGCAGCAAATCGAATGGCTGTCC 1883

    109 lnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGly 125
        ||:::  ||||||:::      |||             |||      :::
   1884 AAACTGGTGGTCATCAAGGTGGGCAGTGT...GCGACCCATCGCAGATCT 1930

    126 GlnIle............ValProThrLeuHisLeuTyrGlnAspLeuLy 138
        :::|||            :::|||:::   ||||||::::::||||||::
   1931 CGTATCCAGGCAGCCCAACTTCCTGCCACCCATCTGCACAAAGATCTCCG 1980

    138 sSerArg.....................GlyIleLeuProAlaAsnThrG 148
        :||||||                    |||   ||||||   ::::::
   1981 GTCGAGAAATAGTAAAATGCAGCTTTCTGGGACCCTTCCTTTCCGTGTCG 2030

    148 lnIleArg.....................SerValTyrArgAsnProGlu 157
          :::|||                    :::||||||::::::   :::
   2031 CTGTTCGCTGAGGAGAACATCAAGTTTGCGGAGTTTACCACCAAGAACAA 2080

    158 LeuAsnGlnCys..............AlaGlyGlyAlaAlaMetSerLy 169
              :::|||               |||   |||||||||::::::
   2081 GCTTGAAGATGCATCCAGCTCTCGATTGCGATGGGAGCTGCACTCAATGC 2130

    169 sHis 170
          |||
   2131 GCAC 2134

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL15779

seq_documentation_block:
ID   ABL15779 standard; cDNA; 5005 BP.
XX
AC   ABL15779;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 41819.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
```

```
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB71676.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 41819; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5005 BP; 1344 A; 1395 C; 1225 G; 1041 T; 0 other;


alignment_scores:
            Quality:   89.50              Length:    263
              Ratio:  0.699                Gaps:     13
 Percent Similarity:  48.669   Percent Identity: 21.673

alignment_block:
 US-09-674-779-2 x ABL15779   ..

 Align seg 1/1  to: ABL15779  from: 1  to: 5005

      23 ProIleProThrAsnProGlnValSerProIle.........LysThrPr 36
         ||||||   ::::::   |||   ||||||:::          ||||||||
    1079 CCCATAAAAAACTCGAGCCAGCCATCACCATTGCCCAAGCCCAAGACACC 1128

      36 oSerValLeuIleThrLysAspLysIleGly......AspHisHisThrH 51
         |:::::      ||||||:::|||            |||      ::::
    1129 GACTATCAAGAGCACCAAGGAGAAGGCCCGCTCGCTGGACTCGGCTGCCA 1178

      51 isGluHisAspGluSer...ValSerHisValGlyLeuGlnAlaHis... 65
         ::|||   :::   |||   |||   |||         ::::::|||
    1179 ACGAATCGGAACTGTCAATCGTGGTGCACAACATAACCGAATCTCATGGC 1228

      66 ............PheGluThrTrpLeuGlnMetHisHisAlaThrLysGl 78
                   :::|||||||      |||   :::   :::|||
    1229 AGTTGTGATGACATGGAGACAGGA...CAGAGCCAGCCATCGACA..... 1270

      78 nGluValValArgTyrGlnAlaTyrLeuGln..................... 88
          :::|||:::::::::::   :::||||||
    1271 .AGTGTTATCCACCATCACCTTCATCTTCAACAGCCCGCATCCGCTATCC 1319

      89 ..............SerArgLeuGlyAsnTyrLeuProProMetSerGln 100
                     ||||||||   :::::|||         :::|||
    1320 ACTCAACATCGCTCTCGCGACTGGATAGTCACCTAAGCTTGGCCAATCAA 1369

     101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGly.............. 112
             |||      |||:::      :::   :::
    1370 AGACGAACGCCGCGCAGAAAGTCACCAGGAATCAACACCACTGACTGGCT 1419

     113 ................HisGluProTyrGlnLeuProProGluHisLeuT 124
                        ::::::||||||||:::   |||   |||
    1420 TATGTACCACCGCAAGCAGAACCCATATCAGGTGCAACCGACACACTGCA 1469

     124 rp............................GlyGlnIleValProThr 130
                                      :::::::::::|||:::
    1470 GCTCCACGACACAAAGCTCCTTGGACTCGGACGCCAGTTTAACCCCCAGT 1519

     131 LeuHisLeuTyrGln.....................AspLeuLysSerAr 140
         |||      ::::::                     |||      |||::
    1520 TTGGGGGACTTTGAACTGAAGTCCGCCTGTAGCGTGGACGGTGGCTCGAA 1569

     140 gGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProG 157
         :      :::   |||:::         ||||||:::::::::::::   :
    1570 GTTTGGAATAGGCGCAAGCCTAGCGCCGAGAAGCGCCCACAAGCACAATC 1619

     157 lu................................LeuAsnGlnCys... 161
         ::                              |||   :::|||
    1620 AGTTACTGCACTCGTCCAGCACAAATTTAAAAACGCTTCCCGAGTGCCTT 1669

     162 .....................AlaGlyGlyAlaAlaMetSerLysHisLe 171
                              :::|||||||:::      ::::::   ::
    1670 ACCCTGGTGGAGTTCTCCTCCTCTGGCGGTGGTCCTAAGGAATCGCCTTT 1719

     171 uThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerG 188
         :   :::::::::::::|||:::   :::|||   |||:::   |||:::
    1720 CAAGCAGAAGTCAATGGATTTGCCGATGCCGACACTCCAGGCAAAGACCA 1769

     188 lnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHis 204
           :::      :::   :::|||   |||                ::::::
    1770 CAGTGTCTACATCATCGATGAATCTCCTT.............CAGCGTCGG 1807

     205 GlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThr 217
         |||:::|||::::::::::   |||:::|||::::::::::
    1808 GGATCCAATCACAGTTTAACGCTCAACCTGCACAGCTCT 1846

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL15778

seq_documentation_block:
ID   ABL15778 standard; cDNA; 7062 BP.
XX
AC   ABL15778;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 41816.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB71675.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 41816; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
```

```
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7062 BP; 1662 A; 1658 C; 1831 G; 1911 T; 0 other;


alignment_scores:
            Quality:   89.50              Length:    263
              Ratio:  0.699                Gaps:     13
 Percent Similarity:  48.669   Percent Identity: 21.673

alignment_block:
 US-09-674-779-2 x ABL15778/rev   ..

 Align seg 1/1  to reverse of: ABL15778  from: 1  to: 7062

      23 ProIleProThrAsnProGlnValSerProIle.........LysThrPr 36
         ||||||   ::::::   |||   |||||||:::          ||||||||
    4984 CCCATAAAAAACTCGAGCCAGCCATCACCATTGCCCAAGCCCAAGACACC 4935

      36 oSerValLeuIleThrLysAspLysIleGly......AspHisHisThrH 51
         |:::::      ||||||:::|||             |||      ::::
    4934 GACTATCAAGAGCACCAAGGAGAAGGCCCGCTCGCTGGACTCGGCTGCCA 4885

      51 isGluHisAspGluSer...ValSerHisValGlyLeuGlnAlaHis... 65
         ::|||   :::   |||   |||   |||          ::::::|||
    4884 ACGAATCGGAACTGTCAATCGTGGTGCACAACATAACCGAATCTCATGGC 4835

      66 ............PheGluThrTrpLeuGlnMetHisHisAlaThrLysGl 78
                    :::|||||||      |||   :::   :::|||
    4834 AGTTGTGATGACATGGAGACAGGA...CAGAGCCAGCCATCGACA..... 4793

      78 nGluValValArgTyrGlnAlaTyrLeuGln.................... 88
          :::|||:::::::::::::   :::||||||
    4792 .AGTGTTATCCACCATCACCTTCATCTTCAACAGCCCGCATCCGCTATCC 4744

      89 ..............SerArgLeuGlyAsnTyrLeuProProMetSerGln 100
                     |||||||||   ::::::|||          :::|||
    4743 ACTCAACATCGCTCTCGCGACTGGATAGTCACCTAAGCTTGGCCAATCAA 4694

     101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGly.............. 112
             |||      |||:::      :::   :::
    4693 AGACGAACGCCGCGCAGAAAGTCACCAGGAATCAACACCACTGACTGGCT 4644

     113 ................HisGluProTyrGlnLeuProProGluHisLeuT 124
                        ::::::||||||||||:::   |||   |||
    4643 TATGTACCACCGCAAGCAGAACCCATATCAGGTGCAACCGACACACTGCA 4594

     124 rp............................GlyGlnIleValProThr 130
                                       ::::::::::::|||:::
    4593 GCTCCACGACACAAAGCTCCTTGGACTCGGACGCCAGTTTAACCCCCAGT 4544

     131 LeuHisLeuTyrGln.....................AspLeuLysSerAr 140
         |||      ::::::                    |||      |||::
    4543 TTGGGGGACTTTGAACTGAAGTCCGCCTGTAGCGTGGACGGTGGCTCGAA 4494

     140 gGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProG 157
         :      :::   |||:::         ||||||:::::::::::::   :
    4493 GTTTGGAATAGGCGCAAGCCTAGCGCCGAGAAGCGCCCACAAGCACAATC 4444

     157 lu................................LeuAsnGlnCys... 161
         ::                              |||   :::|||
    4443 AGTTACTGCACTCGTCCAGCACAAATTTAAAAACGCTTCCCGAGTGCCTT 4394

     162 ....................AlaGlyGlyAlaAlaMetSerLysHisLe 171
                            :::|||||||:::     ::::::   ::
    4393 ACCCTGGTGGAGTTCTCCTCCTCTGGCGGTGGTCCTAAGGAATCGCCTTT 4344

     171 uThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerG 188
         :   ::::::::::::|||:::   :::|||   |||:::   |||:::

    4343 CAAGCAGAAGTCAATGGATTTGCCGATGCCGACACTCCAGGCAAAGACCA 4294

     188 lnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHis 204
           :::      :::   :::|||   |||                ::::::
    4293 CAGTGTCTACATCATCGATGAATCTCCTT.............CAGCGTCGG 4256

     205 GlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThr 217
         |||:::|||:::::::::   |||:::|||::::::::
    4255 GGATCCAATCACAGTTTAACGCTCAACCTGCACAGCTCT 4217
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS35758

seq_documentation_block:

```
ID   AAS35758 standard; DNA; 19206 BP.
XX
AC   AAS35758;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human cardiovascular system antigen genomic DNA SEQ ID No 1258.
XX
KW   Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; tissue regeneration;
KW   anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155321-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01340.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
```

PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451930/48.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
PS Claim 1; SEQ ID No 1258; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 89.50 | Length: | 183 |
| Ratio: | 1.041 | Gaps: | 10 |
| Percent Similarity: | 46.995 | Percent Identity: | 25.683 |

```
alignment_block:
 US-09-674-779-2 x AAS35758    ..

 Align seg 1/1  to: AAS35758  from: 1  to: 19206

      10 ThrThrLeuIleSerSerMetLeuValAlaCysSerAla..... 22
         |||   ::::::|||      |||:::|||||||||:::
    8870 ACCTGTTTTGTCTCACTGGGGCTCCTAGCATGCTCTTCCTTGGCTGGGTG 8919

      23 .....ProIleProThrA 27
                 |||:::|||   :
    8920 TGGGCTCTGCCTTTTCTGCCCGCAAAGCCCTGGGGTGCCGGTACCCTGTT 8969

      27 snProGlnValSerProIleLysThrProSerValLeuIleThr..... 41
         ::   :::   ||||||          |||::::::   |||
    8970 CCTTTTCACCATCTCCCCAGGTGATCTCATCCATCGTCCACACCATGAGC 9019

      42 ...LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVa 57
            :::      ::::::|||   :::   |||:::            ::
    9020 CCACAGCCAGGCCTCTCTGATCCATACCTGCATCAC.....CT 9057

      57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH 74
         :||||||:::::|||||||||      :::|||
    9058 TTCACACATGTCCCTTCAGGCGGCCCCTCAGACT..... 9091

      74 isAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg 90
                                 ||||||   :::||||||||||
    9092 .........................CAAGCT...GTCCAGAGTAGA 9109

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSe 107
         :::|||   :::   ||||||:::::::|||:::||||||
    9110 GTTGGC...CATGGCCCACCATTTACCCAGATGCTCACACCTGAAACTCG 9156

     107 rTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuT 124
               :::      |||      :::|||||||||
    9157 AGTTCTTTGCTCCCACCACATCCATCACCAGCTCCCA..... 9193

     124 rpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
            :::::::::   :::|||   |||   |||           :::
    9194 .....TCGGTTCTCTAGTCTCTGCCCCTAGTGCAA.....CTGCAA 9229

     141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl 157
         ::::::||||||::::::
    9230 TCCTTCTTGCCCTCGTCTCCTGTG..... 9253

     157 uLeuAsnGlnCys...AlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
               ||||||   |||   ::::::::::::|||   |||||||||
    9254 .......CAGTGCCTAGCTCAGGCCACCACTGTCTCAGTGCATCTCACA 9295

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS35754

seq_documentation_block:
ID   AAS35754 standard; DNA; 19274 BP.
XX
AC   AAS35754;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human cardiovascular system antigen genomic DNA SEQ ID No 1254.
XX
KW   Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; tissue regeneration;
KW   anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155321-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01340.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
```

PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451930/48.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
PS Claim 1; SEQ ID No 1254; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

alignment_scores:
Quality: 89.50 Length: 183
Ratio: 1.041 Gaps: 10
Percent Similarity: 46.995 Percent Identity: 25.683

alignment_block:
US-09-674-779-2 x AAS35754 ..

Align seg 1/1 to: AAS35754 from: 1 to: 19274

```
      10 ThrThrLeuIleSerSerMetLeuValAlaCysSerAla..... 22
         |||   :::::::|||      |||:::||||||||||:::
    8870 ACCTGTTTTGTCTCACTGGGGCTCCTAGCATGCTCTTCCTTGGCTGGGTG 8919

      23 .....ProIleProThrA 27
                          |||:::|||   :
    8920 TGGGCTCTGCCTTTTCTGCCCGCAAAGCCCTGGGGTGCCGGTACCCTGTT 8969

      27 snProGlnValSerProIleLysThrProSerValLeuIleThr..... 41
         ::   :::   ||||||           |||::::::   |||
    8970 CCTTTTCACCATCTCCCCAGGTGATCTCATCCATCGTCCACACCATGAGC 9019

      42 ...LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVa 57
            :::      ::::::|||   :::   |||:::          ::
    9020 CCACAGCCAGGCCTCTCTGATCCATACCTGCATCAC.....CT 9057

      57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH 74
         :||||||::::::|||||||||      :::|||
    9058 TTCACACATGTCCCTTCAGGCGGCCCCTCAGACT..... 9091

      74 isAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg 90
                                  ||||||   :::||||||||||
    9092 .....CAAGCT...GTCCAGAGTAGA 9109

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSe 107
         :::|||   :::   ||||||::::::|||:::||||||
    9110 GTTGGC...CATGGCCCACCATTTACCCAGATGCTCACACCTGAAACTCG 9156

     107 rTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuT 124
```

```
              :::      |||      :::|||||||||
    9157 AGTTCTTTGCTCCCACCACATCCATCACCAGCTCCCA..... 9193

     124 rpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
          :::::::::   :::|||   |||   |||           :::
    9194 .....TCGGTTCTCTAGTCTCTGCCCCTAGTGCAA.........CTGCAA 9229

     141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl 157
         ::::::||||||::::::
    9230 TCCTTCTTGCCCTCGTCTCCTGTG..... 9253

     157 uLeuAsnGlnCys...AlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
              ||||||   |||   :::::::::::|||   |||||||||
    9254 .......CAGTGCCTAGCTCAGGCCACCACTGTCTCAGTGCATCTCACA 9295

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS35759

seq_documentation_block:
ID   AAS35759 standard; DNA; 31348 BP.
XX
AC   AAS35759;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human cardiovascular system antigen genomic DNA SEQ ID No 1259.
XX
KW   Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; tissue regeneration;
KW   anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155321-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01340.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
```

```
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451930/48.
XX
PT   New cardiovascular system related polynucleotides and polypeptides,
PT   useful for diagnosing, treating and/or preventing disorders of the
PT   cardiovascular system -
XX
PS   Claim 1; SEQ ID No 1259; 674pp; English.
XX
CC   Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC   the cardiovascular system antigen polypeptides of the invention.
CC   Cardiovascular system antigens and their associated polynucleotides are
CC   useful in the diagnosis, treatment and prevention of various types of
CC   disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC   chickens or sheep. A pathological condition can be determined by
CC   detecting the presence or absence of a mutation in a cardiovascular
CC   system antigen polynucleotide. The treatable disorders include autoimmune
CC   diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC   as neoplasms of the breast or liver, cardiovascular disorders such as
CC   cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC   nervous system disorders such as Alzheimer's disease, infections caused
CC   by bacteria, viruses and fungi, ocular disorders such as corneal
CC   infection, endocrine disorders such as premature labour and infertility,
CC   gastrointestinal disorders such as Crohn's disease, renal disorders such
CC   as glomerulonephritis and respiratory disorders such as asthma and
CC   pleurisy. The polypeptides can also be used to aid wound healing, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, to regenerate tissues and in chemotaxis.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
alignment_scores:
            Quality:   89.50              Length:    183
              Ratio:  1.041                Gaps:     10
 Percent Similarity:  46.995   Percent Identity: 25.683

alignment_block:
 US-09-674-779-2 x AAS35759   ..

 Align seg 1/1  to: AAS35759  from: 1  to: 31348

      10 ThrThrLeuIleSerSerMetLeuValAlaCysSerAla..... 22
         |||   :::::::|||      |||:::||||||||||:::
    8870 ACCTGTTTTGTCTCACTGGGGCTCCTAGCATGCTCTTCCTTGGCTGGGTG 8919

      23 .....................................ProIleProThrA 27
                                                |||:::|||   :
    8920 TGGGCTCTGCCTTTTCTGCCCGCAAAGCCCTGGGGTGCCGGTACCCTGTT 8969

      27 snProGlnValSerProIleLysThrProSerValLeuIleThr..... 41
         ::.  :::   ||||||           |||::::::   |||
    8970 CCTTTTCACCATCTCCCCAGGTGATCTCATCCATCGTCCACACCATGAGC 9019

      42 ...LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVa 57
            :::      ::::::|||   :::   |||:::            ::
    9020 CCACAGCCAGGCCTCTCTGATCCATACCTGCATCAC...........CT 9057

      57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH 74
         :||||||::::::|||||||||      :::|||
    9058 TTCACACATGTCCCTTCAGGCGGCCCCTCAGACT.............. 9091

      74 isAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg 90
                                      ||||||   :::||||||||||
    9092 ............................CAAGCT...GTCCAGAGTAGA 9109

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSe 107
         :::|||   :::   ||||||::::::|||:::||||||
    9110 GTTGGC...CATGGCCCACCATTTACCCAGATGCTCACACCTGAAACTCG 9156

     107 rTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuT 124
               :::      |||      :::|||||||||
    9157 AGTTCTTTGCTCCCACCACATCCATCACCAGCTCCCA............. 9193

     124 rpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
              :::::::::   :::|||   |||   |||            :::
    9194 .....TCGGTTCTCTAGTCTCTGCCCCTAGTGCAA.........CTGCAA 9229

     141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl 157
         ::::::||||||::::::
    9230 TCCTTCTTGCCCTCGTCTCCTGTG........................... 9253

     157 uLeuAsnGlnCys...AlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
               ||||||   |||   ::::::::::::|||   |||||||||
    9254 .......CAGTGCCTAGCTCAGGCCACCACTGTCTCAGTGCATCTCACA 9295

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK79227

seq_documentation_block:
ID   AAK79227 standard; DNA; 31348 BP.
XX
AC   AAK79227;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34039.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
```

XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX

```
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 34039; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 31348 BP; 7245 A; 7873 C; 8030 G; 8200 T; 0 other;


alignment_scores:
            Quality:   89.50              Length:    183
              Ratio:  1.041                Gaps:     10
 Percent Similarity:  46.995   Percent Identity: 25.683

alignment_block:
 US-09-674-779-2 x AAK79227   ..

 Align seg 1/1  to: AAK79227  from: 1  to: 31348

      10 ThrThrLeuIleSerSerMetLeuValAlaCysSerAla..... 22
         |||   :::::::|||       |||:::||||||||||:::
    8870 ACCTGTTTTGTCTCACTGGGGCTCCTAGCATGCTCTTCCTTGGCTGGGTG 8919

      23 .....ProIleProThrA 27
                          |||:::|||   :
    8920 TGGGCTCTGCCTTTTCTGCCCGCAAAGCCCTGGGGTGCCGGTACCCTGTT 8969

      27 snProGlnValSerProIleLysThrProSerValLeuIleThr..... 41
         ::   :::   ||||||           |||::::::   |||
    8970 CCTTTTCACCATCTCCCCAGGTGATCTCATCCATCGTCCACACCATGAGC 9019

      42 ...LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVa 57
            :::      :::::::|||   :::   |||:::          ::
    9020 CCACAGCCAGGCCTCTCTGATCCATACCTGCATCAC.....CT 9057

      57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH 74
         :||||||:::::::|||||||||      :::|||
    9058 TTCACACATGTCCCTTCAGGCGGCCCCTCAGACT..... 9091

      74 isAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg 90
                                |||||||   :::||||||||||
    9092 .........................CAAGCT...GTCCAGAGTAGA 9109

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSe 107
         :::|||   :::   ||||||:::::::|||:::|||||||
    9110 GTTGGC...CATGGCCCACCATTTACCCAGATGCTCACACCTGAAACTCG 9156

     107 rTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuT 124
               :::      |||      :::||||||||||
    9157 AGTTCTTTGCTCCCACCACATCCATCACCAGCTCCCA..... 9193

     124 rpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
            :::::::::   :::|||   |||   |||           :::
    9194 .....TCGGTTCTCTAGTCTCTGCCCCTAGTGCAA.....CTGCAA 9229

     141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl 157
         ::::::||||||::::::
    9230 TCCTTCTTGCCCTCGTCTCCTGTG..... 9253

     157 uLeuAsnGlnCys...AlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
                 ||||||   |||   ::::::::::::|||   |||||||||
    9254 .......CAGTGCCTAGCTCAGGCCACCACTGTCTCAGTGCATCTCACA 9295

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAD10354

seq_documentation_block:
ID   AAD10354 standard; cDNA; 2479 BP.
XX
AC   AAD10354;
XX
DT   24-SEP-2001  (first entry)
XX
DE   Mouse tumour necrosis factor receptor/osteoprotegerin-like cDNA.
XX
KW   Mouse; cytostatic; tumour necrosis factor receptor/osteoprotegerin;
KW   TNFr/OPG-like protein; antisense inhibitor; gene therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             91..1401
FT                   /*tag=  a
FT                   /product= "Mouse TNFr/OPG protein"
FT   sig_peptide     91..183
FT                   /*tag=  b
FT   mat_peptide     184..1398
FT                   /*tag=  c
FT                   /product= "Mouse mature TNFr/OPG protein"
XX
PN   WO200144472-A1.
XX
PD   21-JUN-2001.
XX
PF   14-DEC-2000; 2000WO-US33858.
XX
PR   16-DEC-1999;   99US-0172306.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Jing S,  Welcher AA,  Fox GM,  Shu J,  Boedigheimer MJ,  Bennett BD;
XX
DR   WPI; 2001-451665/48.
DR   P-PSDB; AAE05518.
XX
PT   New tumor necrosis factor receptor / osteoprotegerin-like
PT   (TNFr/OPG-like) polypeptides useful for diagnosis and treatment of
PT   associated disease -
XX
PS   Claim 1; Fig 2; 208pp; English.
XX
CC   The present sequence is a cDNA encoding mouse tumour necrosis factor
CC   receptor/osteoprotegerin (TNFr/OPG)-like protein. TNFr/OPG-like protein
CC   is useful for the treatment, prevention or amelioration of a medical
CC   condition in a mammal resulting from decreased levels of TNFr/OPG
CC   protein. Nucleic acid sequences of the present invention are used to
CC   map locations of the TNFr/OPG genes and related genes. They are also
CC   used as antisense inhibitors of TNFr/OPG expression. TNFr/OPG-like
CC   protein and cDNA sequences are also used in gene therapy.
XX
SQ   Sequence 2479 BP; 498 A; 711 C; 725 G; 545 T; 0 other;


alignment_scores:
            Quality:   89.00              Length:    188
              Ratio:  1.035                Gaps:     10
 Percent Similarity:  45.745   Percent Identity: 21.809

alignment_block:
```

```
US-09-674-779-2 x AAD10354/rev   ..

Align seg 1/1  to reverse of: AAD10354  from: 1  to: 2479

     18 ValAlaCysSerAlaProIleProThrAsnProGlnVal..... 30
        :::||||||   :::|||:::||||||   |||::::::
   2212 CTAGCCTGCCTGTCTCCTCTTCCCACAGCTCCAGAGCTTGAGGGCAGCAG 2163

     31 .......................SerProIleLysThrProSerVal. 38
                               ||||||:::::::::   |||
   2162 GGTGGACTGAGGCTCTGCACCAGAGAGCCCTGTCCAAGTCAAGTCATTTT 2113

     39 .....LeuIleThrLysAspLysIle.GlyAspHisHisThr........ 50
           :::|||   :::::::::::: |||:::   |||:::
   2112 TCAGTTTCATTAAAAATAGCAGGGTAGGGGCAGGCACATGCATTAAACCC 2063

     51 ..........HisGluHisAspGluSerVal......SerHisValGlyL 62
                 |||:::|||:::::::::         ||||||   |||
   2062 CTTCTGAAGGCATCACCATAATAGACAGTCCCATGGGAGCCATGAAGGC. 2014

     62 euGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGln 78
          |||:::      ::::::      :::         ||||||:::
   2013 ..CAAGGCCCAGGAAGCAGCAAAAACAGGCCAGATAAAGCTACTAATGGG 1966

     79 GluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsn..... 93
        :::|||::::::  ::::::      |||:::          |||
   1965 AGGGTAGCAAAGTCAAGCTCCCCCCAACAGGACTCAGCCCAGAATCTGCA 1916

     93 ................................................. 93

   1915 AAGAGCTGGAGAGGTCCTGGGAACTGAGTTCTCCTCCCCTACGTACCTAG 1866

     94 .......................TyrLeuProProMetSerGlnLeu 101
                              |||           :::   |||
   1865 GGATAAGCCCAGGATTGGAACCCAGGTACTCTGCCACCAGTACAGGGCTC 1816

    102 LeuThrThrAlaArgSerTrpGlnAlaCys..................... 111
        ::::::|||         ::::::||||||
   1815 TTTGCTACCCCTTCCCTCTATAGGGCCTGCCAGAAGAGAGGGAGTCACAT 1766

    112 ..........GlyHis...GluProTyrGlnLeuProProGluHisLeuT 124
                |||||   :::   :::   :::|||      :::   |
   1765 CCAGGGTCCAGGACATGGGCAAGAGTGGCTAGTTCCCAAGGCACAAGCTT 1716

    124 rpGlyGlnIle...ValProThrLeuHisLeuTyrGlnAspLeuLysSer 139
        |||||||||:::   :::|||   ||||||      :::   |||::::::
   1715 GGGGCCAGGTACAGATACCACTCCTCCACCAAACCAGACCTCTAAGGGAC 1666

    140 ArgGlyIleLeu 143
        ||||||:::|||
   1665 AGAGGGCTCCTA 1654

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL02149

seq_documentation_block:
ID   ABL02149 standard; cDNA; 1362 BP.
XX
AC   ABL02149;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 929.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB58046.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 929; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1362 BP; 319 A; 399 C; 375 G; 269 T; 0 other;


alignment_scores:
            Quality:   88.50            Length:    226
              Ratio:  0.903              Gaps:      9
 Percent Similarity:  43.363   Percent Identity: 21.681

alignment_block:
 US-09-674-779-2 x ABL02149   ..

 Align seg 1/1  to: ABL02149  from: 1  to: 1362

      32 ProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHi 48
         ||||||   |||   |||   |||   :::::::|||          :::
     129 CCTATATACACT...TCTTTTTTGACATCGCGAGACGGAGCCAAAAATGC 175

      48 sHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaH 65
                    |||   ::::::   :::|||::::::          |
     176 TGGAATAATCGGCCATGGCAATGAGAGCGCTCATCTTTCTGGCCCTCGCC 225

      65 isPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValVal 81
         ||       |||   :::       :::             |||
     226 ACTCTGGTGGCTGGCGAAGGTCTTCGGCTTCCGGACCAGCAGTCGTCCAA 275

      82 ArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMe 98
         ::::::   |||   |||::::::          :::          |||
     276 CAACATCCAGCAAGTCTACGCACCGCAGCCACCGGCTCAGCAGATCCAGC 325

      98 tSerGlnLeuLeuThrThr.................AlaArgSerTrpG 109
          :::          ||||||                ||||||||||||:
     326 AACCCCAGCAGCAACAACAGTTTCCACAGGCACAGCAGCCAGGAGTTGGC 375

     109 lnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrp... 124
         :::::                        ||||||   :::|||||||||
     376 GAGGA.........ACGCGGACGCTCCTCCCTGCTCAGCATCTTTGGCCT 416

     125 GlyGlnIleValProThrLeu............................. 131
         ||||||         :::|||
     417 GGGCAACGACAACGATCCCTTCCTGGCCAGGACCAACAGCAACTGCCTCG 466

     132 ....................................HisLeuTyrG 135
                                             ||||||   :
     467 GCGGCGACCTCTCCGAGTGCTTCAAGACTCAGGCGCTCAACACCTTCGAC 516
```

```
      135 lnAspLeuLysSerArgGlyIleLeuProAlaAsn...ThrGlnIleArg 150
          ::||||||   :::      :::       :::       :::
      517 GAGATCTTCTTCAAGGACCAGTACAAGCTCTCCGACTTTGCTCGTGTGGT 566

      151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGly.......... 163
          ::::::   |||||||||           :::|||||||
      567 GCGACTGCCAGAAACCCAGCAGAGGTCCTTGCTGCAGGAGCCATTCGAGT 616

      164 .......GlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspI 178
                |||||||:::      ::::::   ::::::|||||   :::|
      617 ACTCCGAGGAGCCACGTGGCGACGACGACGAGTGGAATCAGCTGCTGAAA 666

      178 leTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGln 194
          |||||:::|||            |||:::::::::|||:::::::::
      667 TATGGACTCCG...........AAGGGCCGAACGCTTCATCAAGTC... 701

      195 AsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGl 211
                                |||||||::::::   :::   ::
      702 .......................CACGGCTCTAGAGGTCGAGTGGC 724

      211 yLeuGlyLeuTyrAlaThrGlyAlaIle 220
          :   |||   :::   :::   ::::::
      725 CCGAGGAACTCACCGAAGCCGGTCGTTA 752
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS30562

seq_documentation_block:

ID AAS30562 standard; DNA; 1580 BP.
XX
AC AAS30562;
XX
DT 21-NOV-2001 (first entry)
XX
DE DNA encoding novel prostate gland antigen, Seq ID No 420.
XX
KW Human; nootropic; neuroprotective; cytostatic; antiparkinsonian;
KW antianaemic; dermatological; immunosuppressive; antiinflammatory;
KW antiarthritic; antirheumatic; virucide; hepatotropic; nephrotropic;
KW osteopathic; prostate gland; prostatitis; adenocarcinoma; hair loss;
KW prostatosis; malacoplakia; adenocarcinoma; benign prostatic hypertrophy;
KW hyperplasia; carcinoma; prostate neoplastic disorder; skin aging;
KW reproductive system disorder; autoimmune disorder; urinary system;
KW systemic lupus erythematosus; rheumatoid arthritis; cardiovascular;
KW blood-related disorder; hyperproliferative disorder; respiratory;
KW neurological disorder; endocrine disorder; inflammatory disorder;
KW liver disorder; wound healing; food preservative; ds.
XX
OS Homo sapiens.
XX
PN WO200155447-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01330.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

```
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476223/51.
XX
PT   Novel isolated prostate gland related polypeptide useful for diagnosis
PT   and treatment of disorders of prostate such as prostatodystonia,
PT   prostatosis, prostatitis, benign prostatic hypertrophy and malacoplakia
PT   -
XX
PS   Claim 1; SEQ ID No 420; 512pp; English.
XX
CC   The invention relates to novel isolated prostate gland related nucleic
CC   acids (I) and polypeptides (II). (I) and (II) are useful for diagnosis,
CC   prognosis, prevention, and/or treatment of diseases and/or disorders of
CC   the prostate such as acute non-bacterial prostatitis, chronic non-
CC   bacterial prostatitis, acute bacterial prostatitis, prostatodystonia,
CC   prostatosis, granulomatous prostatitis, malacoplakia, benign prostatic
CC   hypertrophy or hyperplasia, and prostate neoplastic disorders, including
CC   adenocarcinomas, transitional cell carcinomas, ductal carcinomas, and
CC   squamous cell carcinomas. (I), (II) and antibody to (II) are useful for
CC   diagnosing and treating reproductive system disorders (Paget's disease),
CC   autoimmune disorders (systemic lupus erythematosus, rheumatoid
CC   arthritis), blood-related disorders (sickle cell anaemia),
CC   hyperproliferative disorders, urinary system disorders
CC   (glomerulonephritis), cardiovascular disorders (arrhythmias),
CC   respiratory disorders, musculoskeletal system disorders, neural activity
CC   and neurological disorders (Alzheimer's disease and Parkinson's disease),
CC   endocrine disorders (Addison's disease), gastrointestinal disorders
CC   (inflammatory disorders), liver disorders (biliary liver cirrhosis),
CC   pancreatic and gall bladder disorders, disorders of the large intestine,
CC   developmental and inherited disorders, diseases at the cellular level,
CC   and wound healing and epithelial cell proliferation. (I) or (II) is
CC   useful to prevent skin aging, for preventing hair loss, to maintain

alignment_scores:
           Quality:   88.50            Length:    267
             Ratio:  0.770              Gaps:     15
 Percent Similarity:  43.071   Percent Identity: 22.846

alignment_block:
 US-09-674-779-2 x AAS30562   ..

 Align seg 1/1  to: AAS30562  from: 1  to: 1580

      20 CysSerAlaProIleProThrAsnProGlnVal.SerProIleLysThrP 36
         |||      |||   |||||||:::   |||::: |||   :::
     236 TGCCCCTTACCCCCACCAACACAAACACAGGCAGTCAGATTTATGGCGAA 285

      36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHis... 51
           ::::::   ||||||:::       :::      ||||||
     286 TGGCCATTTCTATAACAAATGCAACGGTGACCAAGCACCACCCTCCCGCA 335

      52 ......GluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPh 66
              :::|||   :::   ::::::   :::|||||||||
     336 ATGCTAAGGCACCCACAGAGGGCAGGCATAGCTGGTCTCCAAATTTTT.. 383

      66 eGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgT 83

     383 ................................................. 383

      83 yrGlnAlaTyrLeuGlnSerArg...LeuGlyAsnTyrLeuProProMet 98
                 |||:::||||||   |||   :::||||||      :::
     384 ........GCATTACACTCAAGGGTCCTGGAGTCATATCTAGAAAAGGTC 425

      99 SerGlnLeuLeuThrThrAlaArg.........SerTrpGlnAlaCysGl 112
         ::::::           ||||||    :::   ::::::  ||
     426 CAGATGGCTTCAATCCTAGCCAGACCCACAGCCACACCTTCATCCTTGGG 475

     112 yHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValP 129
         ||||:::      :::|||||||   :::   ||||||
     476 ACACAGCTCTCTAAAACTCCCATATGACATCCTCTGG.............. 512

     129 roThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla 145
           :::|||:::|||   |||
     513 ..AGCCTCAGACTCCTTCAGGGA........................... 533

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
                 :::   :::::::      :::|||   |||          ||
     534 ......CCTGTCTGAGGCCTCCAACCACACCCTCTGCTG.........GC 568

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
         |||||||   |||      :::      :::   |||   |||   |
     569 AGGTGGGCCTGCAGGAAGGGAGGGAGCACCTGGGCCAGCCTGTGATCGGT 618

     179 rpVal.............................................. 180
         ||
     619 GGTGTCTGAGTCCCTCTGGTCTGGCTTAGAGAGAGCATCCTGCCTTCACC 668

     181 ...............................ProAspLeuGluIleLy 186
                                        ::::::::::::
     669 ACCTCAGGCTGGAGAGTGAAGACACGGTAGCCCCTGAGTATGTCTGTACC 718

     186 sSerGln.AlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeu 202
           |||::: :::||||||:::   ::::::      :::|||   :::
     719 AAGCAGGGGTGCTCTACAGTAGATCCGAGGCCAGAGCCCAGTCCTATACA 768

     203 GluHisGlyGluAsnGlnAsnPheGlyLeuGly.....LeuTy 215
```

```
      |||:::|||   :::::::::|||   :::|||              ::
  769 GAGAGAGGAGGTAGAAGGTCATTCCAGGTGGGACCCAGAAGAGTTCCTTT 818

  215 rAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyA 232
      ::::      |||:::|||   ::::::|||          |||||||||
  819 TTCTTGCCCCGCCCTGCACTGGAATGCCCAG.........AAATGGGGTT 859

  232 laGlnPheSerGluThrAsnSerIleCysArgHisValLeuProLys 247
          |||   |||      |||:::|||   |||   ||||||:::
  860 TA...TTTCACGAG......TCCTTCTGTCCCCAC...CTACCACGA 894

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL06616

seq_documentation_block:
ID   AAL06616 standard; DNA; 1580 BP.
XX
AC   AAL06616;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9304.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
```

```
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition  -
XX
PS   Disclosure; SEQ ID NO 9304; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 1580 BP; 353 A; 467 C; 401 G; 359 T; 0 other;


alignment_scores:
            Quality:   88.50              Length:    267
              Ratio:  0.770                Gaps:     15
 Percent Similarity:  43.071   Percent Identity: 22.846

alignment_block:
 US-09-674-779-2 x AAL06616   ..

 Align seg 1/1  to: AAL06616  from: 1  to: 1580

      20 CysSerAlaProIleProThrAsnProGlnVal.SerProIleLysThrP 36
         |||      |||   ||||||:::   |||::: |||   :::
     236 TGCCCCTTACCCCCACCAACACAAACACAGGCAGTCAGATTTATGGCGAA 285

      36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHis... 51
           ::::::   |||||||:::      :::       ||||||
     286 TGGCCATTTCTATAACAAATGCAACGGTGACCAAGCACCACCCTCCCGCA 335

      52 ......GluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPh 66
              :::|||   :::   ::::::   :::||||||||||
     336 ATGCTAAGGCACCCACAGAGGGCAGGCATAGCTGGTCTCCAAATTTTT.. 383

      66 eGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgT 83

     383 ................................................. 383

      83 yrGlnAlaTyrLeuGlnSerArg...LeuGlyAsnTyrLeuProProMet 98
                   |||:::||||||   |||   :::||||||      :::
     384 ........GCATTACACTCAAGGGTCCTGGAGTCATATCTAGAAAAGGTC 425

      99 SerGlnLeuLeuThrThrAlaArg.........SerTrpGlnAlaCysGl 112
         ::::::              ||||||        :::   ::::::   ||
     426 CAGATGGCTTCAATCCTAGCCAGACCCACAGCCACACCTTCATCCTTGGG 475

     112 yHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValP 129
         ||||:::      :::||||||   :::   ||||||
     476 ACACAGCTCTCTAAAACTCCCATATGACATCCTCTGG.............. 512

     129 roThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla 145
           :::|||:::|||   |||
     513 ..AGCCTCAGACTCCTTCAGGGA.......................... 533

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
                 :::   ::::::     :::|||   |||         ||
     534 ......CCTGTCTGAGGCCTCCAACCACACCCTCTGCTG.........GC 568

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
         |||||||   |||      :::       :::   |||   |||   |
     569 AGGTGGGCCTGCAGGAAGGGAGGGAGCACCTGGGCCAGCCTGTGATCGGT 618

     179 rpVal.............................................. 180
         ||
     619 GGTGTCTGAGTCCCTCTGGTCTGGCTTAGAGAGAGCATCCTGCCTTCACC 668

     181 ...............................ProAspLeuGluIleLy 186
                                        ::::::::::::
     669 ACCTCAGGCTGGAGAGTGAAGACACGGTAGCCCCTGAGTATGTCTGTACC 718

     186 sSerGln.AlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeu 202
          |||::: :::||||||:::   ::::::     :::|||   :::
     719 AAGCAGGGGTGCTCTACAGTAGATCCGAGGCCAGAGCCCAGTCCTATACA 768

     203 GluHisGlyGluAsnGlnAsnPheGlyLeuGly.............LeuTy 215
         |||:::|||   ::::::::::|||   :::|||             ::
     769 GAGAGAGGAGGTAGAAGGTCATTCCAGGTGGGACCCAGAAGAGTTCCTTT 818

     215 rAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyA 232
         ::::      |||:::|||   ::::::|||         |||||||||
     819 TTCTTGCCCCGCCCTGCACTGGAATGCCCAG.........AAATGGGGTT 859

     232 laGlnPheSerGluThrAsnSerIleCysArgHisValLeuProLys 247
           |||   |||      |||:::|||   |||   |||||||:::
     860 TA...TTTCACGAG......TCCTTCTGTCCCCAC...CTACCACGA 894


seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS80217

seq_documentation_block:
ID   AAS80217 standard; cDNA; 900 BP.
XX
AC   AAS80217;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #16021.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
```

```
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG16030.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 16021; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 900 BP; 218 A; 262 C; 218 G; 202 T; 0 other;


alignment_scores:
            Quality:   88.00              Length:    294
              Ratio:  0.638                Gaps:     12
 Percent Similarity:  46.939   Percent Identity: 20.068

alignment_block:
 US-09-674-779-2 x AAS80217    ..

 Align seg 1/1  to: AAS80217  from: 1  to: 900

       8 PheIleThrThrLeuIleSerSerMetLeuValAla..............  19
         :::::::::|||   :::::::::   ::::::|||
      61 TATGTGGCCACTGCAGTAGCAGCACGCATCATAGCAGATGTTCAAGGAAT 110

      20 ..........CysSerAlaProIleProThrAsnProGlnValSerProI  33
                   |||             |||:::   |||:::   :::   :
     111 CTCTGATACTTGTGTGATTGCGCAGCCTTCACCACCACACCCTGGAGCTT 160

      33 leLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHis  49
         :::::|||      :::::::::|||   |||:::
     161 TGAGGACC......TTAGTATTCACCACAGATGAA.............. 189

      50 ThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPh  66
            |||      :::      :::   :::|||:::|||
     190 ...CCGGAGACTCCTCAGCCCCCAGGTGCAGCAGGATTCCAGCTGGTCCA 236

      66 eGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgT  83
         :::      :::|||:::|||::::::           |||:::
     237 GAAGAAACAAGTGCAGCTGCATAACGTACTT......TATGTAATGACAC 280

      83 yrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSer  99
           :::|||:::   |||:::   :::      |||||||||      :::
     281 CCCATGCCTTTGCGCAAGGAACGGTGACTATTTATCTGCCTGGCGAACAA 330

     100 GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTy 116
         |||   |||::::::|||          ::::::
     331 CAGACACTTTCCGTTGCCCCCTTGAAAAATGTTGTCCAGTTGGTGACACA 380

     116 rGlnLeuProProGluHisLeuTrp..........................  124
          |||         ::::::||||||
     381 ACAGCATCTGCGCGATCGCCTCTGGTGGCCTGGTGCCTTCCTGACCGACT 430

     125 ..GlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg  140
           :::      |||   :::|||      :::      ::::::   :::
     431 TTGCGGCGAAAGTTAAAGCGCTTAAAGACTTCCCAACCATGTCATGGCAC 480

     141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl  157
         ::::::||||||::::::   :::      :::::::::|||:::   ::
     481 AGCGTGCTTCCGGGGAAGCAGAAGCCGACGACGATGTGGCGGCAA...CG 527

     157 uLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThr....  172
         :||||||   ||||||::::::   :::::::|||      |||:::
     528 ATTAAATCCGTGCGCCAGCAGTTACTCAATCTCAATATCACTGGTCGCTT 577

     173 .....................................AsnSerAlaIle  176
                                              ::::::||||||
     578 ACCGGTCAAACTGGACCCTGATTTTGTACGAGTGGATGAGAACAGCAATC 627

     177 AspIleTrpVal.....................................  180
             |||
     628 CCCCGTTGGTCGGGGACTATACGCTCTATACCGTACAGCGTCCGGGTACC 677

     181 ............................ProAspLeuGlu...IleLysS  187
                                      |||:::||||||   :::|||
     678 GAGCAAGGGTCGTTCAAGGTGAAGCGGGCCCGAGCTAGAGTTCGTCAAAC 727

     187 erGlnAlaLeuTyrGluLeuGlnAsn...ArgLeuCysGlnTyrTrpLeu  202
               ||||||      |||:::   :::|||   :::   |||
     728 CCGGGCCCCTTTATTTGGACCAAAGTCCCCAGCTCCTGAGCACATGG... 774

     203 GluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAl  219
               :::::::|||:::|||::::::|||:::      |||:::
     775 ......AGTCAGAACTCGAATTTGAGCTTATCGGCCCAAGCAAATCAACA 818

     219 aIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnPheSerG  236
          :::   :::      :::||||||         |||         |||
     819 AGTGGGAATC...CTAGAGGGCTTCGTTCTGAAAGGCCAGTTCTCTTCTG 865

     236 luThrAsnSerIleCysArgHisValLeuPro  246
               ::::::   |||      :::|||||||
     866 CCTGTACCCAGAGATGCCCACTGCTGCTGCCA 897

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH81475

seq_documentation_block:
ID   AAH81475 standard; DNA; 963 BP.
XX
AC   AAH81475;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Escherichia coli protein encoding nucleotide sequence SEQ ID NO:274.
XX
KW   Escherichia coli; identification; proliferation; microorganism;
KW   antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW   bacterial growth inhibition; ds.
XX
OS   Escherichia coli.
XX
PN   WO200148209-A2.
XX
PD   05-JUL-2001.
XX
PF   19-DEC-2000; 2000WO-US34419.
XX
PR   23-DEC-1999;   99US-0173005.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen KL,  Zyskind JW;
```

```
XX
DR   WPI; 2001-457376/49.
DR   P-PSDB; AAG98419.
XX
PT   Novel nucleic acids encoding proteins required for Escherichia coli
PT   proliferation, useful for screening for antimicrobial agents -
XX
PS   Example 4; Page 411-412; 596pp; English.
XX
CC   The present invention describes a purified or isolated nucleic acid
CC   sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC   given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC   microorganism is capable of inhibiting proliferation of a microorganism.
CC   (I) have antibacterial and antibiotic activities, and can be used in
CC   gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC   of the microorganism, and the manufactured antibiotic is useful for
CC   reducing the activity or level of a gene product required for
CC   proliferation of a microorganism in a subject, specifically humans. The
CC   nucleic acids that inhibit bacterial growth or proliferation can be used
CC   as antisense therapeutics for killing bacteria. In addition to
CC   therapeutic applications, the nucleic acid sequences complementary to
CC   sequences required for proliferation can be used as diagnostic tools.
CC   For example, nucleic acid probes complementary to proliferation-required
CC   sequences that are specific for particular species of microorganisms can
CC   be used as probes to identify particular microorganism species in
CC   clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC   proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC   represent oligonucleotides, which are used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 963 BP; 232 A; 257 C; 256 G; 218 T; 0 other;


alignment_scores:
            Quality:   88.00              Length:    255
              Ratio:  0.807                Gaps:     11
 Percent Similarity:  42.745   Percent Identity: 20.784

alignment_block:
 US-09-674-779-2 x AAH81475   ..

 Align seg 1/1  to: AAH81475  from: 1  to: 963

       8 PheIleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIl 24
         ::::::|||   :::      :::      ||||||   :::   |||
      28 TGGCTAACGTTTTTTACGTTCGCCGCTGCCGTGGCGCTGGCGCTACCGGC 77

      24 eProThrAsnProGlnValSerProIleLysThrProSerValLeuIleT 41
          :::|||           |||      :::   |||   |||:::
      78 AAAAGCCAACACCTGGCCGCTGCCG...CCAGCGGGCAGTCGTCTGGTTG 124

      41 hrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57
         ::::::|||:::         |||:::   |||:::|||   :::
     125 GCGAAAACAAATTT.........CATGTGGTGGAAAATGACGGTGGT... 162

      58 SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHi 74
                                          :::::::::
     163 ....................................TCTCTGGAAGC 173

      74 sAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgL 91
          :::|||:::      |||   :::   |||   ||||||:::::
     174 CATCGCCAAAAAATACAACGTCGGCTTTCTCGCTCTGTTACAGGCTAACC 223

      91 euGly......AsnTyrLeuProProMetSerGlnLeuLeuThrThrAla 105
         |||         |||:::|||      :::::::::||||||
     224 CCGGCGTTGATCCTTACGTACCGCGCGCGGGCAGCGTGTTAACG...... 267

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHi 122

     267 ..................................................  267

     122 sLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysS 139
                        :::|||         |||      |||

     268 ................ATCCCGTTGCAAACCCTACTTCCAGATGCGCCGC 301

     139 erArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsn 155
           :::|||||||:::         :::::::::|||      |||
     302 GCGAAGGCATTGTGATCAACATTGCGGAGCTGCGTCTCTATTACTACCCG 351

     156 ProGluLeuAsnGlnCysAla.........................Gl 163
         |||      |||:::   :::                          ||
     352 CCGGGTAAAAATTCGGTAACCGTGTATCCAATAGGTATTGGTCAGTTAGG 401

     163 yGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp....... 177
         ||||   :::::::::      :::::::::::::::   |||
     402 TGGTGACACGCTGACACCGACAATGGTGACCACCGTTTCAGACAAACGTG 451

     178 ........IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr 191
                   |||:::|||      :::|||:::      |||   |||
     452 CAAACCCAACCTGGACGCCAACGGCAAACATCCGC......GCCCGTTAT 495

     192 GluLeuGlnAsn..................................... 195
         :::   |||:::
     496 AAAGCACAGGGAATTGAGTTGCCTGCGGTAGTGCCGGCTGGACTGGATAA 545

     196 .....................ArgLeuCysGlnTyr.........TrpL 202
                               ||||||:::   |||      :::|
     546 CCCAATGGGCCATCATGCGATTCGTCTGGCGGCCTATGGCGGCGTTTATT 595

     202 euGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu...TyrAlaThr 217
         ||   ||||||   |||   :::||||||:::|||:::      ::::::
     596 TGCTTCATGGTACGAACGCCGATTTCGGCATTGGCATGCGGGTAAGTTCT 645

     218 GlyAlaIleHisLeu 222
         |||:::|||:::|||
     646 GGCTGTATTCGTCTG 660

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH84500

seq_documentation_block:
ID   AAH84500 standard; DNA; 963 BP.
XX
AC   AAH84500;
XX
DT   26-SEP-2001  (first entry)
XX
DE   E. coli growth and proliferation related coding sequence SEQ ID NO:128.
XX
KW   Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW   bacterial infection; microorganism; ds.
XX
OS   Escherichia coli.
XX
PN   WO200134810-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30950.
XX
PR   09-NOV-1999;   99US-0164415.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen K,  Zyskind J;
XX
DR   WPI; 2001-335933/35.
DR   P-PSDB; AAG99078.
XX
PT   Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT   for screening for homologous genes and for designing expression vectors
PT   -
XX
PS   Claim 9; Page 130-132; 522pp; English.
XX
CC   AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC   related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
```

```
CC   growth and proliferation related proteins given in AAG99078 and AAG98830
CC   to AAG98999. (I) can be used as potential targets for the generation of
CC   new antimicrobial agents, and for identification of compounds which
CC   interact with the gene products of (I). In addition the expression of
CC   (I) and the purification of the proteins, the purified proteins can be
CC   used to generate reagents and screen small molecule libraries or other
CC   candidate compound libraries for compounds that can be further developed
CC   to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC   complementary to (I) that are specific for particular species of
CC   microorganisms can be used to identify particular microorganism species
CC   in clinical specimens, therefore, providing a rapid and dependable
CC   method by which to identify the causative agents of a bacterial
CC   infection. Also, antibodies generated against proteins translated from
CC   mRNA transcribed from proliferation-required sequences can also be used
CC   to screen for specific microorganisms that produce such proteins in a
CC   species-specific manner. AAH84371 and AAH84670 represent sequencing
CC   primers used in the isolation of E. coli growth and proliferation
CC   related sequence, which are used in an example from the present
CC   invention.
XX
SQ   Sequence 963 BP; 232 A; 257 C; 256 G; 218 T; 0 other;


alignment_scores:
            Quality:   88.00              Length:    255
              Ratio:  0.807                Gaps:     11
 Percent Similarity:  42.745   Percent Identity: 20.784

alignment_block:
 US-09-674-779-2 x AAH84500  ..

 Align seg 1/1  to: AAH84500  from: 1  to: 963

       8 PheIleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIl 24
         :::::::|||   :::       :::     ||||||   :::   |||
      28 TGGCTAACGTTTTTTACGTTCGCCGCTGCCGTGGCGCTGGCGCTACCGGC 77

      24 eProThrAsnProGlnValSerProIleLysThrProSerValLeuIleT 41
          :::|||          |||      :::   |||   |||:::
      78 AAAAGCCAACACCTGGCCGCTGCCG...CCAGCGGGCAGTCGTCTGGTTG 124

      41 hrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57
         :::::::|||:::         |||:::   |||:::|||   :::
     125 GCGAAAACAAATTT.........CATGTGGTGGAAAATGACGGTGGT... 162

      58 SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHi 74
                                          :::::::::
     163 .....................................TCTCTGGAAGC 173

      74 sAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgL 91
          :::|||:::      |||   :::   |||   ||||||::::::
     174 CATCGCCAAAAAATACAACGTCGGCTTTCTCGCTCTGTTACAGGCTAACC 223

      91 euGly......AsnTyrLeuProProMetSerGlnLeuLeuThrThrAla 105
         |||         |||:::|||      ::::::::::||||||
     224 CCGGCGTTGATCCTTACGTACCGCGCGCGGGCAGCGTGTTAACG...... 267

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHi 122

     267 ............................................... 267

     122 sLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysS 139
                    :::|||         |||      |||
     268 ..............ATCCCGTTGCAAACCCTACTTCCAGATGCGCCGC 301

     139 erArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsn 155
         :::||||||:::         :::::::::|||     |||
     302 GCGAAGGCATTGTGATCAACATTGCGGAGCTGCGTCTCTATTACTACCCG 351

     156 ProGluLeuAsnGlnCysAla.........................Gl 163
         |||      |||:::   :::                          ||
     352 CCGGGTAAAAATTCGGTAACCGTGTATCCAATAGGTATTGGTCAGTTAGG 401

     163 yGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp....... 177
         ||||   :::::::::      :::::::::::::::   |||
     402 TGGTGACACGCTGACACCGACAATGGTGACCACCGTTTCAGACAAACGTG 451

     178 ........IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr 191
                 |||:::|||      :::|||:::      |||   |||
     452 CAAACCCAACCTGGACGCCAACGGCAAACATCCGC......GCCCGTTAT 495

     192 GluLeuGlnAsn..................................... 195
         :::   |||:::
     496 AAAGCACAGGGAATTGAGTTGCCTGCGGTAGTGCCGGCTGGACTGGATAA 545

     196 .....................ArgLeuCysGlnTyr.........TrpL 202
                               ||||||:::   |||         :::|
     546 CCCAATGGGCCATCATGCGATTCGTCTGGCGGCCTATGGCGGCGTTTATT 595

     202 euGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu...TyrAlaThr 217
         ||   ||||||   |||   :::|||||||:::|||:::      ::::::
     596 TGCTTCATGGTACGAACGCCGATTTCGGCATTGGCATGCGGGTAAGTTCT 645

     218 GlyAlaIleHisLeu 222
         |||:::|||:::|||
     646 GGCTGTATTCGTCTG 660

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS90352

seq_documentation_block:
ID   AAS90352 standard; cDNA; 977 BP.
XX
AC   AAS90352;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #26156.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG26165.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 26156; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
```

```
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 977 BP; 240 A; 277 C; 237 G; 223 T; 0 other;


alignment_scores:
            Quality:   88.00              Length:    294
              Ratio:  0.638                Gaps:     12
 Percent Similarity:  46.939   Percent Identity: 20.068

alignment_block:
 US-09-674-779-2 x AAS90352   ..

 Align seg 1/1  to: AAS90352  from: 1  to: 977

       8 PheIleThrThrLeuIleSerSerMetLeuValAla..............  19
         ::::::::::|||   :::::::::   :::::::|||
      61 TATGTGGCCACTGCAGTAGCAGCACGCATCATAGCAGATGTTCAAGGAAT 110

      20 ..........CysSerAlaProIleProThrAsnProGlnValSerProI  33
                   |||           |||:::   |||:::   :::   :
     111 CTCTGATACTTGTGTGATTGCGCAGCCTTCACCACCACACCCTGGAGCTT 160

      33 leLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHis  49
         :::::|||      ::::::::|||   |||:::
     161 TGAGGACC......TTAGTATTCACCACAGATGAA.............. 189

      50 ThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPh  66
            |||      :::      :::   :::|||:::|||
     190 ...CCGGAGACTCCTCAGCCCCCAGGTGCAGCAGGATTCCAGCTGGTCCA 236

      66 eGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgT  83
       :::      :::|||:::|||::::::            |||:::
     237 GAAGAAACAAGTGCAGCTGCATAACGTACTT......TATGTAATGACAC 280

      83 yrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSer  99
         :::|||:::   |||:::   :::      |||||||||      :::
     281 CCCATGCCTTTGCGCAAGGAACGGTGACTATTTATCTGCCTGGCGAACAA 330

     100 GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTy 116
         |||   |||:::::::|||         ::::::
     331 CAGACACTTTCCGTTGCCCCCTTGAAAAATGTTGTCCAGTTGGTGACACA 380

     116 rGlnLeuProProGluHisLeuTrp........................  124
          |||         ::::::||||||
     381 ACAGCATCTGCGCGATCGCCTCTGGTGGCCTGGTGCCTTCCTGACCGACT 430

     125 ..GlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
         :::      |||   :::|||      :::      ::::::   :::
     431 TTGCGGCGAAAGTTAAAGCGCTTAAAGACTTCCCAACCATGTCATGGCAC 480

     141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl 157
         ::::::|||||::::::   :::      ::::::::::|||:::   ::
     481 AGCGTGCTTCCGGGGAAGCAGAAGCCGACGACGATGTGGCGGCAA...CG 527

     157 uLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThr.... 172
         :||||||   ||||||::::::   ::::::|||      |||:::
     528 ATTAAATCCGTGCGCCAGCAGTTACTCAATCTCAATATCACTGGTCGCTT 577

     173 .....................................AsnSerAlaIle 176
                                             ::::::||||||
     578 ACCGGTCAAACTGGACCCTGATTTTGTACGAGTGGATGAGAACAGCAATC 627

     177 AspIleTrpVal..................................... 180
         |||
     628 CCCCGTTGGTCGGGGACTATACGCTCTATACCGTACAGCGTCCGGGTACC 677

     181 ..........................ProAspLeuGlu...IleLysS 187
                                    |||:::||||||   :::|||
     678 GAGCAAGGGTCGTTCAAGGTGAAGCGGGCCCGAGCTAGAGTTCGTCAAAC 727

     187 erGlnAlaLeuTyrGluLeuGlnAsn...ArgLeuCysGlnTyrTrpLeu 202
               ||||||      |||:::   :::|||   :::   |||
     728 CCGGGCCCCTTTATTTGGACCAAAGTCCCCAGCTCCTGAGCACATGG... 774

     203 GluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAl 219
         :::::::|||:::|||::::::|||:::      |||:::
     775 ......AGTCAGAACTCGAATTTGAGCTTATCGGCCCAAGCAAATCAACA 818

     219 aIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnPheSerG 236
          :::   :::      :::||||||         |||         |||
     819 AGTGGGAATC...CTAGAGGGCTTCGTTCTGAAAGGCCAGTTCTCTTCTG 865

     236 luThrAsnSerIleCysArgHisValLeuPro 246
         ::::::   |||      :::|||||||
     866 CCTGTACCCAGAGATGCCCACTGCTGCTGCCA 897

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL03309

seq_documentation_block:
ID   ABL03309 standard; cDNA; 1014 BP.
XX
AC   ABL03309;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 4409.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59206.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 4409; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence 1014 BP; 283 A; 234 C; 232 G; 265 T; 0 other;


alignment_scores:
           Quality:   88.00              Length:    211
             Ratio:  0.936                Gaps:     13
 Percent Similarity:  44.550   Percent Identity: 27.962

alignment_block:
 US-09-674-779-2 x ABL03309   ..

 Align seg 1/1  to: ABL03309  from: 1  to: 1014

      48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
         ||||||   ::::::|||:::          :::|||
     402 CATCATATACGCCATCATGAAGGGAGATGCCAGCAC.............. 437

      64 aHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValV 81
                    |||   ::::::||||||   ::::::   :::   |
     438 .............TGG...AGTTTCCATCATGAAAATCGAACCTCACCGG 471

      81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
         |||||||||:::   |||::::::|||      ::::::|||      |||
     472 TTCGATATCGGTGATATTTTAGCTCAGCGAGAAGTGGCTATTA.ATCCCG 520

      98 MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
         ||||||                  |||   :::   |||   |||
     521 ATGTCT.....................TCATGCCGGATCTGCATGCATCT 549

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrL 131
               :::      |||::::::||||||                   |
     550 TTGGCGTCGCTAGGAGCCGAGTTACTTGTGG............ATACTAT 587

     131 euHisLeuTyrGln...AspLeuLysSerArgGlyIleLeuProAlaAsn 146
         ||      ||||||   ||||||::::::
     588 TAACAACCTATCAGTAAGACTTAAGGAAG..................... 616

     147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
                              ||||||::::::   |||
     617 .....................CCAAACCCCAAGATAGTTCAA........ 637

     163 yGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpV 180
          |||:::||||||                          :
     638 .GGGTCAGCTATG.................................C 650

     180 alProAspLeuGluIleLysSerGlnAlaLeu.................. 190
         ::|||   |||   :::||||||||||   |||
     651 TCCCAAGATTACCAGTAAAATCACAGAAATTATCTGGTCCGAGCTGAGTG 700

     191 .........TyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHi 204
                 |||   |||       :::||||||      ||||||:::
     701 CCATTGAAATATACACTCGTCATCGAGCTTTGTA.....TGGCTACAAAA 745

     204 sGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGly...AlaI 220
               ||||||   ||||||   :::      :::::::|||   ||||
     746 ACCTTACAACCAACTTTTTGGGCAAGCAAGTGCAGTTGCTGGAGTTGCGA 795

     220 leHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnPheSerGlu 236
         ||                              |||   :::   |||:::
     796 TT..........................ACCGGAGAAGAGAGTTCCCGT 818

     237 ThrAsnSerIleCysArgHisValLeuProLys 247
            ::::::   |||   |||   |||||||||
     819 CCAAGAACCTGGTGC...CATCAGCTACCTAAG 848

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS46271

seq_documentation_block:
ID   AAS46271 standard; DNA; 12368 BP.
XX
AC   AAS46271;
XX
DT   18-DEC-2001  (first entry)
XX
DE   DNA encoding novel mar regulated protein (NIMR) #40.
XX
KW   mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
OS   Escherichia coli.
XX
PN   WO200170776-A2.
XX
PD   27-SEP-2001.
XX
PF   08-MAR-2001; 2001WO-US07478.
XX
PR   10-MAR-2000; 2000US-188362P.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Levy SB,  Barbosa TM,  Alekshun MN;
XX
DR   WPI; 2001-602769/68.
DR   P-PSDB; AAU29372.
XX
PT   Identifying compounds that modulate a newly identified mar regulated
PT   polypeptide activity, useful as antimicrobial compounds, involves
PT   contacting the polypeptide with a test compound -
XX
PS   Disclosure; Page 458-466; 526pp; English.
XX
CC   The invention relates to a method of identifying compounds that modulate
CC   a newly identified mar regulated (NIMR) polypeptide activity. The method
CC   comprises contacting an NIMR polypeptide with a test compound under
CC   interaction conditions, determining the ability of the compound to
CC   modulate the activity or expression of the polypeptide, and selecting the
CC   modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC   of microbial infections, and in screening for modulators of NIMR
CC   expression and activity. These modulators can be used to reduce the
CC   infectivity of a microbe on a surface, and the virulence of a microbe in
CC   a subject suffering from an infection. AAS46232-AAS46278 represent
CC   Escherichia coli NIMR coding sequences of the invention.
XX
SQ   Sequence 12368 BP; 2959 A; 3148 C; 3276 G; 2985 T; 0 other;


alignment_scores:
           Quality:   88.00              Length:    255
             Ratio:  0.807                Gaps:     11
 Percent Similarity:  42.745   Percent Identity: 20.784

alignment_block:
 US-09-674-779-2 x AAS46271/rev   ..

 Align seg 1/1  to reverse of: AAS46271  from: 1  to: 12368

       8 PheIleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIl 24
         ::::::|||   :::      :::      ||||||   :::   |||
    8659 TGGCTAACGTTTTTTACGTTCGCCGCTGCCGTGGCGCTGGCGCTACCGGC 8610

      24 eProThrAsnProGlnValSerProIleLysThrProSerValLeuIleT 41
            :::|||            |||      :::   |||   |||:::
    8609 AAAAGCCAACACCTGGCCGCTGCCG...CCAGCGGGCAGTCGTCTGGTTG 8563

      41 hrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57
         :::::::|||:::         |||:::   |||:::|||   :::
    8562 GCGAAAACAAATTT.........CATGTGGTGGAAAATGACGGTGGT... 8525

      58 SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHi 74
                                                 :::::::::
    8524 .....................................TCTCTGGAAGC 8514

      74 sAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgL 91
            :::|||:::      |||   :::   |||   ||||||:::::
    8513 CATCGCCAAAAAATACAACGTCGGCTTTCTCGCTCTGTTACAGGCTAACC 8464
```

```
     91 euGly......AsnTyrLeuProProMetSerGlnLeuLeuThrThrAla 105
          |||         |||:::|||      :::::::::||||||
   8463 CCGGCGTTGATCCTTACGTACCGCGCGCGGGCAGCGTGTTAACG...... 8420

    106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHi 122

   8420 ..................................................... 8420

    122 sLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysS 139
                      :::|||         |||       |||
   8419 ..............ATCCCGTTGCAAACCCTACTTCCAGATGCGCCGC 8386

    139 erArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsn 155
          :::||||||:::         :::::::::|||      |||
   8385 GCGAAGGCATTGTGATCAACATTGCGGAGCTGCGTCTCTATTACTACCCG 8336

    156 ProGluLeuAsnGlnCysAla.........................Gl 163
          |||      |||:::   :::                       ||
   8335 CCGGGTAAAAATTCGGTAACCGTGTATCCAATAGGTATTGGTCAGTTAGG 8286

    163 yGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp..... 177
          ||||   :::::::::      :::::::::::::::   |||
   8285 TGGTGACACGCTGACACCGACAATGGTGACCACCGTTTCAGACAAACGTG 8236

    178 ........IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr 191
                  |||:::|||      :::|||:::       |||   |||
   8235 CAAACCCAACCTGGACGCCAACGGCAAACATCCGC......GCCCGTTAT 8192

    192 GluLeuGlnAsn..................................... 195
          :::   |||:::
   8191 AAAGCACAGGGAATTGAGTTGCCTGCGGTAGTGCCGGCTGGACTGGATAA 8142

    196 .....................ArgLeuCysGlnTyr.........TrpL 202
                              |||||||:::   |||         :::|
   8141 CCCAATGGGCCATCATGCGATTCGTCTGGCGGCCTATGGCGGCGTTTATT 8092

    202 euGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu...TyrAlaThr 217
          ||   ||||||   |||   :::||||||:::|||:::      ::::::
   8091 TGCTTCATGGTACGAACGCCGATTTCGGCATTGGCATGCGGGTAAGTTCT 8042

    218 GlyAlaIleHisLeu 222
          |||:::|||:::|||
   8041 GGCTGTATTCGTCTG 8027
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA82625

seq_documentation_block:

```
ID   ABA82625 standard; DNA; 66933 BP.
XX
AC   ABA82625;
XX
DT   25-JAN-2002  (first entry)
XX
DE   Human HBM gene region b200e21-h_contig4.
XX
KW   Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW   sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW   antisense therapy; vaccine; bone disorder; Paget's disease;
KW   sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177327-A1.
XX
PD   18-OCT-2001.
XX
PF   21-JUN-2000; 2000WO-US16951.
XX
PR   05-APR-2000; 2000US-0543771.
PR   05-APR-2000; 2000US-0544398.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Carulli JP,  Little RD,  Recker RR,  Johnson ML;
XX
DR   WPI; 2001-657171/75.
XX
PT   New high bone mass (HBM) and Zmax1 genes and proteins useful for
PT   modulating bone mass for the treatment of e.g. osteoporosis -
XX
PS   Claim 51; Page 308-350; 443pp; English.
XX
CC   The present invention describes the human Zmax1 gene and the high bone
CC   mass (HBM) gene,which are found on chromosome 11q13.3. The Zmax1 and
CC   HBM genes have osteopathic activities. The genes can be used in gene
CC   therapy, antisense therapy and in the production of vaccines. They
CC   can be used in the diagnosis and treatment of bone disorders including
CC   osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC   dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;
```

alignment_scores:

```
           Quality:   88.00          .   Length:    248
             Ratio:  0.880               Gaps:     15
 Percent Similarity:  40.323   Percent Identity: 26.210
```

alignment_block:

US-09-674-779-2 x ABA82625/rev ..

Align seg 1/1 to reverse of: ABA82625 from: 1 to: 66933

```
     23 ProIleProThrAsnProGlnVal..............SerProIleLy 34
          |||      |||:::||||||:::              |||||||||
  39344 CCCACACACACACACCCCCAGATGGAAAGGGTGCCTCCCTCCCCCATACA 39295

     34 sThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
          |||    :::                              ||||||||||
  39294 CACACACACACCA.....................CACACACACCACACAC 39266

     51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
          ||   |||   ::::::   :::|||:::       |||
  39265 ACACACACACACACACACACACACACACACACAGCCTTGCT......... 39225

     68 ThrTrpLeuGlnMetHisHis............AlaThrLysGlnGluVa 80
              ||||||   ||||||                ||||||::::::||
  39224 GGGCAGCTGCAAAGACATCATCACTTTCATCTGGACACGAAGGATCATGT 39175

     80 lValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProP 97
          |
  39174 G..................................................... 39174

     97 roMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGly... 112
              |||      :::|||||||:::   |||   |||:::||||||
  39173 .....TCTTTGTAGATAACAACATCCACATCACTACAGAGCTGTGGACCC 39129

    113 .....................HisGluProTyrGlnLeuProProGluHi 122
                             :::                |||||   :::||
  39128 TAGGGGGCCATGTGTCAAAGCCAACTGAACAAATTCCTTCCTAAAAGCCA 39079

    122 s...LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuL 138
          |   ||||||               ||||||      |||:::
  39078 CTAGCTGTGG..............ACCCTGGACTGGTACAGC....... 39051

    138 ysSerArgGlyIleLeuProAlaAsnThrGlnIleArg.SerValTyrAr 154
                             |||:::   |||   ||| |||   |||
  39050 ..............TGCCGGGCTGAAAGACAAGCAAGATTCCAGCTAC.. 39017

    154 gAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
          |||      ||||||:::      |||:::
  39016 ....CCTGGCCAAAATCAGGCTCACCGTGGAGGG.............. 38987

    171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
```

```
                                |||||||          ::::::::::::
    38986 .....................TGGGTCTGTGGGAACCAGATGCAAAAA 38960

      188 GlnAlaLeuTyrGluLeuGlnAsnArg.....................Le 197
          |||:::|||   :::   :::::::::                     ||
    38959 CAGACTTTGGGACAGCAGTCTGGAGAACCACCATCTGAGAAACGTGTACT 38910

      197 uCysGlnTyr......TrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212
          ||||:::         |||   :::   ::::::   |||:::
    38909 GTGTGACGTGACACCCTGGCCACACTCCTCCCGGGCACAAAAACAACACA 38860

      212 euGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg 228
            |||||||||:::|||   :::|||              |||
    38859 GCAAACTTTACGCGGCTGGCCAGTTTCATAAACCCTGAACTCTTTTT... 38813

      229 LysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArg 242
                      |||   ||||||:::|||      |||
    38812 ..............TTTTTTGAGACGGAGTCTTGCGTGAGA 38786
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK94427

seq_documentation_block:

ID AAK94427 standard; cDNA; 1409 BP.
XX
AC AAK94427;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 3204.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR P-PSDB; AAM93499.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 3204; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 1409 BP; 291 A; 460 C; 414 G; 244 T; 0 other;

```
alignment_scores:
           Quality:   87.00              Length:    272
             Ratio:  0.806                Gaps:     14
 Percent Similarity:  39.706   Percent Identity: 22.794

alignment_block:
 US-09-674-779-2 x AAK94427/rev   ..

 Align seg 1/1  to reverse of: AAK94427  from: 1  to: 1409

      25 ProThrAsnProGlnValSerProIleLysThrPro..............  36
         ||||||:::|||   :::   |||      :::|||
    1070 CCCACCCACCCCCTGGCCCTGCCAGCTGGTAGTCCATCAGCACAATGAAG 1021

      37 ............SerValLeuIleThrLysAspLysIleGlyAspHisH  49
                     |||:::|||   :::   :::      ||||||:::
    1020 GAGACTCACATACTCCCTGCTCCCTGTGGTAGAGACACCTGGTGATCAGA 971

      49 isThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHis  65
                                          |||   |||   ||||||
     970 AG.......................TGTCTGGTTTGCTTGGCTGCCCAT 945

      66 PheGlu...ThrTrp......LeuGlnMetHisHisAlaThrLysGlnGl  79
         :::         |||       |||   :::       :::   :::   ::
     944 TTGCCTCTTGAGTGGGCAGCCCTGGGCTTGGGCCCCTCCCTCCGGCCCTC 895

      79 uValValArgTyrGlnAla.............................. 85
         :|||:::      |||
     894 AGTGTTGGCTCTGCAGAAGCTCTGGGGTTCCCTTCAAGTGCACGAGGGGT 845

      86 .................TyrLeuGlnSerArgLeuGlyAsnTyrLeuPro  96
                          |||:::|||   ||||||:::::|||
     844 TAGGCTGCTGTCCCTGAGTCCTCCATTCTGTACTGGGGGGCTGGCTAGGA 795

      97 ProMetSerGlnLeuLeuThrThrAlaArgSer...TrpGlnAla..... 110
         |||   :::              |||   |||   |||||||||
     794 CCTGGGGCTGTGGCCTCTCAGGGGGCAGCCTCTCCATGGCAGGCATCCCT 745

     111 .....................................CysGlyHisGluP 115
                                              |||      |
     744 GCCTTGGGCTGCCCTCCCCCAGACCCCTGACCACCCCCTGGGTCCTGTCC 695

     115 roTyrGlnLeuProProGluHisLeuTrpGlyGlnIle......ValPro 129
         ||:::|||   |||   :::   |||||||||:::          |||
     694 CCCACCAGAGCCCCAGCTCCTGTCTGTGGGGGAGCCATCACGGTGTTCGT 645

     130 ThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAs 146
         :::::::|||      ::::::      |||  |||      |||   ::
     644 GCAGTCCATAGCGCTTCTCAATGTGTGTCACCCGGAACCTGCCCACAGAG 595

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
         :||||||                              :::      :::|
     594 AACACAG.........................TGGTCTCCCCTGTCCG 572

     163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
         |||||:::   :::                                 |||
     571 GGGGCGGCTTTTTCCTTCCT.........................TGG 549

     180 ...ValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAs 195
            |||||||||   ::::::::::::::::::
     548 AGCGTCCCTGACGGACAAGTGGAGGCCTCTTGC................. 516

     195 nArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212
               |||                          :::   |||   |||
     515 .......TGC.......................GGCTGCAATGGAGGCA 497

     212 euGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg 228
            ||||||   :::   |||||||:::      |||   :::|||:::|||
     496 AGGGGCTGCAGAGCCCAGGTGCACTGTGTGATGATGGGAGGGGGCTCCGT 447

     229 .............LysTrpGlyAlaGlnPheSerGluThrAsnSerIleCy 241
                      :::|||      :::::::::   :::   |||:::
     446 CCTGCAGGCTGGAGGTGGCATCCACACTGGACAGCTGCACTGTCCCTTCT 397
```

```
     241 sArgHisValLeuPro 246
             |||      |||
     396 CCTTCACTGTCCCCCA 381

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK80624

seq_documentation_block:
ID   AAK80624 standard; DNA; 5858 BP.
XX
AC   AAK80624;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35436.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
```

```
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 35436; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 5858 BP; 1455 A; 1579 C; 1490 G; 1334 T; 0 other;


alignment_scores:
           Quality:   87.00              Length:    162
             Ratio:  1.338                Gaps:      8
 Percent Similarity:  40.123   Percent Identity: 23.457

alignment_block:
 US-09-674-779-2 x AAK80624    ..

 Align seg 1/1  to: AAK80624  from: 1  to: 5858

      20 CysSerAlaProIlePro.........ThrAsnProGlnValSerProIl 33
         |||:::   |||:::|||         ||||||   |||:::   |||
    2330 TGTCAGTAGCCAGTTCCATTCATTTTGACAAACACACAGGCCCACCCAGC 2379

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
            :::   :::|||                                ||||
    2380 TCTGTCCCAGGCAGTG.............................CACA 2399

      50 hrHisGluHisAspGluSerValSerHisValGly.............. 61
         ||||||||||||:::   |||::::::::::   |||
    2400 CACATGAGCATAGCTAATCCACAAAGCAGCCCGGCTGGGTAAATGGTATT 2449

      61 ............................................... 61

    2450 ATGCTCATTTTACAGAGGAGGAAAATTGAGGTTCAGAGAGAAGCCAAGAC 2499

      62 ....LeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrL 77
             |||   :::|||:::      |||         ||||||
    2500 TTACCTGGGGTCCCATATCCCATGCTGGCAAGTGCCACACCACAAACCTG 2549

      77 ysGlnGluValVal.........ArgTyrGlnAlaTyrLeuGlnSerArg 90
           |||::::::::::         |||   :::::::::         |||
    2550 TCCAAAAACTTACCAGCCAGGGAAGGCTGTCAGTCTTTACCTGGAGGAGA 2599

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeu......ThrThrAl 105
         |||:::   |||         :::|||            :::::::||
    2600 GGTGGTGGTAGTCTTGGGAGCAGGCAGCAGGCAGCTCATGGGGCAGTGGC 2649

     105 aArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu..... 118
         ||||:::|||:::         |||:::|||   ||||||
    2650 AAGAGCCTGGTCTCGGGAACCACACAGACCT...CAGCTCAAATCCAGGC 2696

     119 .........................ProProGlu............ 121
                                     ||||||:::
    2697 TCCATCACTGTGTGACTTTAGAAAAATGACCACCCTCTCTGGGACTCAGT 2746

     122 ...HisLeuTrpGlyGlnIleValProThrLeuHis 132
                |||   :::   :::|||     |||
    2747 TTTCCCACATGGAAGATGAGGATACCAATTTCACAT 2782

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK80625

seq_documentation_block:
ID   AAK80625 standard; DNA; 5866 BP.
XX
AC   AAK80625;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35437.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
```

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 35437; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 5866 BP; 1464 A; 1580 C; 1489 G; 1333 T; 0 other;

```
alignment_scores:
           Quality:   87.00            Length:    162
             Ratio:  1.338              Gaps:      8
 Percent Similarity:  40.123   Percent Identity: 23.457

alignment_block:
 US-09-674-779-2 x AAK80625   ..

 Align seg 1/1  to: AAK80625  from: 1  to: 5866

      20 CysSerAlaProIlePro.........ThrAsnProGlnValSerProIl 33
         |||:::   |||:::|||          ||||||   |||:::   |||
    2329 TGTCAGTAGCCAGTTCCATTCATTTTGACAAACACACAGGCCCACCCAGC 2378

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
           :::   :::|||                                  ||||
    2379 TCTGTCCCAGGCAGTG..............................CACA 2398

      50 hrHisGluHisAspGluSerValSerHisValGly.............. 61
         ||||||||||||:::   |||:::::::::   |||
    2399 CACATGAGCATAGCTAATCCACAAAGCAGCCCGGCTGGGTAAATGGTATT 2448

      61 ............................................... 61

    2449 ATGCTCATTTTACAGAGGAGGAAAATTGAGGTTCAGAGAGAAGCCAAGAC 2498

      62 ....LeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrL 77
             |||   :::|||:::      |||         ||||||
    2499 TTACCTGGGGTCCCATATCCCATGCTGGCAAGTGCCACACCACAAACCTG 2548

      77 ysGlnGluValVal.........ArgTyrGlnAlaTyrLeuGlnSerArg 90
           |||::::::::::         |||   :::::::::         |||
    2549 TCCAAAAACTTACCAGCCAGGGAAGGCTGTCAGTCTTTACCTGGAGGAGA 2598

      91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeu......ThrThrAl 105
         |||:::   |||         :::|||            ::::::||
    2599 GGTGGTGGTAGTCTTGGGAGCAGGCAGCAGGCAGCTCATGGGGCAGTGGC 2648

     105 aArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu.......... 118
         ||||:::|||:::         |||:::|||   ||||||
    2649 AAGAGCCTGGTCTCGGGAACCACACAGACCT...CAGCTCAAATCCAGGC 2695

     119 ............................ProProGlu............ 121
                                     ||||||:::
    2696 TCCATCACTGTGTGACTTTAGAAAAATGACCACCCTCTCTGGGACTCAGT 2745

     122 ...HisLeuTrpGlyGlnIleValProThrLeuHis 132
               |||   :::   :::|||      |||
    2746 TTTCCCACATGGAAGATGAGGATACCAATTTCACAT 2781

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL25734

seq_documentation_block:
ID   ABL25734 standard; DNA; 3864 BP.
XX
AC   ABL25734;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 28675.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 28675; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3864 BP; 1078 A; 764 C; 842 G; 1180 T; 0 other;


alignment_scores:
           Quality:   86.50            Length:    146
             Ratio:  1.153              Gaps:      8
 Percent Similarity:  51.370   Percent Identity: 26.027

alignment_block:
 US-09-674-779-2 x ABL25734/rev   ..

 Align seg 1/1  to reverse of: ABL25734  from: 1  to: 3864

       7 TyrPheIleThrThrLeuIle............SerSerMetLeuValAl 19
         |||:::   ::::::|||:::          |||:::   :::::::||
    1572 TATTACTCCTCAAATCTAGTTTACTTACCCCGAAGTCAACCCATCATAGC 1523

      19 aCysSerAlaProIleProThrAsnPro......GlnValSerProIleL 34
         |   :::         |||:::||||||         |||      :::|
    1522 CATCGAATACTTCCCGCCGGTGAACCCTTTTAAAGTAGTCCATTACCTCA 1473

      34 ysThrProSerVal............LeuIleThrLysAsp......... 43
         ||||||||||||:::           |||:::::::||||||
    1472 AAACTCCTTCTATAAGGATACTGACCCTCGTTTCGAAAGATTACTTCGTA 1423

      44 ......................LysIleGlyAspHisHisThr...Hi 51
                               :::||||||:::   ||||||   ||
    1422 GTCCTGAGGATTCATCGTTAGCACCAAATCGGTTCCTGCCACACTGGGCA 1373

      51 sGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluT 68
         |      |||   |||:::      |||   ::::::   ||||||
    1372 TCCGAAAGATACTTCCATATTGTCGGTTCATATCGAGAAACATTTCGTGG 1323

      68 hrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGln 84
                 :::   ::::::   :::::::::      :::   :::
    1322 ACCGGCAAATTCTTATAGAGACCTTAAGAGAGGATTTCAACAAAAGATCA 1273

      85 AlaTyrLeuGln............SerArgLeuGlyAsnTyrLeuPro.P 97
         :::   |||:::            :::::::|||      :::|||||| |
    1272 ACAATTCTCGAATGAGAGAATACTGGAAAGCTTTTCCTTCACTTACCGCC 1223

      97 roMetSerGlnLeuLeuThrThrAlaArgSerTrp...GlnAlaCysGly 112
         ||   ::::::|||   :::   |||       |||
    1222 CGGCAGAAAAGCTCTGATCAATTGCAATTTACTGGGTCCCGGAATATCAG 1173

     113 HisGluProTyrGlnLeuProProGluHisLeuTrp 124
```

```
      |||:::   :::   ||||||   :::   ||||||
   1172 CATAAGGTTTGGCCTCTTCCAAGTTCACTGCTGTGG 1137

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV18082

seq_documentation_block:
ID   AAV18082 standard; cDNA; 1386 BP.
XX
AC   AAV18082;
XX
DT   17-AUG-1998  (first entry)
XX
DE   Human liver ribonuclease inhibitor cDNA.
XX
KW   Ribonuclease inhibitor; human; nucleic acid; DNA; RNA; polypeptide;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO9806845-A1.
XX
PD   19-FEB-1998.
XX
PF   14-AUG-1997;   97WO-US14254.
XX
PR   13-AUG-1997;   97US-0910731.
PR   16-AUG-1996;   96US-0024057.
PR   03-FEB-1997;   97US-0794546.
PR   04-FEB-1997;   97US-0795395.
XX
PA   (LIFE-) LIFE TECHNOLOGIES INC.
XX
PI   Chatterjee DK,  Shandilya H;
XX
DR   WPI; 1998-159537/14.
DR   P-PSDB; AAW48636.
XX
PT   Mammalian ribonuclease inhibitor genes - useful for developing
PT   products for in vitro production of nucleic acids and polypeptides
XX
PS   Claim 6; Page 49-50; 78pp; English.
XX
CC   This cDNA clone, deposited as NRRL B-21610, codes for human liver
CC   ribonuclease inhibitor (RI) (see AAW48636), a protein useful for the
CC   in vitro production of nucleic acids and polypeptides.  It was
CC   isolated from a human liver cDNA library by PCR amplification (see
CC   AAV18088-89).  A vector comprising the human RI nucleic acid, a host
CC   cell comprising the vector and a method for producing porcine RI by
CC   culturing the host cells are claimed.  RNase-free recombinant RI is
CC   obtained by culturing the host cells and mixing the RI with a
CC   binding partner.  Pig liver, rat liver, human liver and chimeric
CC   pig-liver RIs (see AAW48627-28 and AAW48636-37) are claimed.  The human
CC   and chimeric RI can be used in a claimed method for producing
CC   nucleic acids (including single-stranded or double-stranded DNA,
CC   cDNA, RNA and DNA-RNA hybrids), while pig liver, rat liver, human
CC   liver or the chimeric RI can be used in a claimed method for the in
CC   vitro production of polypeptides.
XX
SQ   Sequence 1386 BP; 249 A; 432 C; 466 G; 239 T; 0 other;


alignment_scores:
           Quality:   86.00            Length:    268
             Ratio:  0.694              Gaps:     15
 Percent Similarity:  46.269   Percent Identity: 25.373

alignment_block:
 US-09-674-779-2 x AAV18082/rev   ..

 Align seg 1/1  to reverse of: AAV18082  from: 1  to: 1386

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||:::   :::   |||   :::|||           ::::::|||::

    1101 CTGGCCCAGGCCCTGGCACAGCTCCCGCACGCCCGCATCCTCCAGCCTGT 1052

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
         :  :::         |||      |||      :::|||:::|||
    1051 TGTTGCTTATCTGTAGCTCCAGGAGAAACCTGTTCTGGGCCAGCAC.... 1006

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
                                 ::::::||||||      :::
    1005 ...................TGAGCTGAAGTGGGGGCAGCAGGCGGCTGT 976

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTy 83
         |||:::         :::|||           |||   ::::::|||::
     975 GAAGCT...GCAGGACTTCACCCACAGCGACTCCAGCTGGCAGCCAGGTT 929

      83 rGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMet.... 98
         :|||      |||  ::::::   |||:::   :::|||   :::
     928 CCAGCAGGGTCTCACACAGCAGTCGGGCACCCTCATCCCCCAGCTCGTTG 879

      99 ........SerGlnLeuLeu.....................ThrThrAla 105
                 ::::::||||||                          ||||||
     878 CCGGCCAGGCTGAGCTCCTTCAGGCTCTCCTTGGCCCTGAGGACACGGCA 829

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro..... 120
         :::          |||   |||   :::             |||
     828 CAGATCCCCGCAGCCCTTGGCAGTGATGCCACACTCCCAGATCCACAGGG 779

     121 ......................GluHisLeuTrpGlyGlnIleValProT 130
                               |||:::   |||:::|||:::
     778 TCCTGAGCCTGGAGCTGGGGTGGAGCAGCCCTGGGCACAGCTC...CGCC 732

     130 hrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsn 146
            |||:::   |||   :::   :::::::|||   ||||||   :::
     731 ATGCCCACATCACCCAGCTTGTTGCTGCCCAGGGCCAGCTCCCGCAGCGA 682

     147 ....................ThrGlnIleArgSerValTyr........ 153
                              :::|||:::   :::|||
     681 GGCCTTGGAGGCCACAATGCCGCACAGGTCCCGGCAGTTGTCTGATGTCA 632

     154 ..........ArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla.... 165
                    :::   |||      |||   ||||||||||:::
     631 CACCGCAGCTCTCCAGCTTGAGCGCCTCCAGCTGGCAGGGGGAGTCCTTC 582

     166 ..AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
           |||::::::::::|||         |||   ||||||:::   |||
     581 AGGCCCTGGCACAGCACACGGACGCCAGCCTCATTGATGTCGTTGTTGCT 532

     182 AspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCy 198
         :::   |||:::                     |||:::   ||||||
     531 AACCGTGAGCTC.....................CTTGAAGTCCGGCTTGG 503

     198 sGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                    |||||||||   :::         |||||:::
     502 CCCT.........GAGCACGGAGGCCAG.........GGGCTCGCAGCTG 471

     215 yrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe.ArgLys...Tr 230
           :::      |||      |||:::   |||   ||| |||:::   ||
     470 GCAGCCGAGAGGCTGCAATACTCCAGCTGCAGCTTTTCCAGGCGGCACTG 421

     230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuProL 247
         ||||::::::   |||   ::::::|||   ||||||   :::   ||||
     420 GGGGTCCAGG...AGTCCTTCGCAGAGCAGCTGCAGGCCCGCATCCCCCA 374

     247 ys 247
         ||
     373 AG 372

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:AAX77394

seq_documentation_block:
ID   AAX77394 standard; DNA; 1614 BP.
XX
AC   AAX77394;
XX
```

```
DT   16-AUG-1999  (first entry)
XX
DE   Human placental ribonuclease inhibitor (PRI) gene sequence.
XX
KW   Human; placental ribonuclease inhibitor; PRI; recombinant; ss.
XX
OS   Homo sapiens.
XX
PN   WO9012881-A.
XX
PD   01-NOV-1990.
XX
PF   18-APR-1990;   90EP-0908084.
XX
PR   24-APR-1989;   89US-0342362.
XX
PA   (PROM-) PROMEGA CORP.
XX
PI   Lewis MK,  Shultz JW,  Schultz JW,  Lewis M;
XX
DR   WPI; 1990-348484/46.
DR   P-PSDB; AAY21644.
XX
PT   Vector contg. gene for human placental ribonuclease inhibitor - used
PT   to express placental ribonuclease inhibitor which is refolded and
PT   purified
XX
PS   Disclosure; Fig 3; 43pp; English.
XX
CC   This represents a cloned gene sequence coding for the human placental
CC   ribonuclease inhibitor (PRI). The invention provides a method whereby the
CC   PRI gene can be cloned invitro using recombinant DNA technology. The
CC   human PRI that is obtained is free from contaminants.
XX
SQ   Sequence 1614 BP; 324 A; 495 C; 504 G; 291 T; 0 other;


alignment_scores:
            Quality:   86.00              Length:    268
              Ratio:  0.694               Gaps:       15
 Percent Similarity:  46.269   Percent Identity: 25.373

alignment_block:
 US-09-674-779-2 x AAX77394/rev   ..

 Align seg 1/1  to reverse of: AAX77394  from: 1  to: 1614

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||:::   :::   |||   :::|||           :::::::|||::
    1116 CTGGCCCAGGCCCTGGCACAGCTCCCGCACGCCCGCATCCTCCAGCCTGT 1067

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
         :   :::         |||      |||      :::|||:::|||
    1066 TGTTGCTTATCTGTAGCTCCAGGAGAAACCTGTTCTGGGCCAGCAC.... 1021

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
                         ::::::||||||      :::
    1020 ...................TGAGCTGAAGTGGGAGCAGCAGGCGGCTGT 991

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTy 83
         |||:::         :::|||           |||   ::::::|||::
     990 GAAGCT...GCAGGACTTCACCCACAGCGACTCCAGCTGGCAGCCAGGTT 944

      83 rGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMet.... 98
         :|||      |||   ::::::   |||:::   :::|||   :::
     943 CCAGCAGGGTCTCACACAGCAGTCGGGGACCCTCATCCCCCAGCTCGTTG 894

      99 ........SerGlnLeuLeu.....................ThrThrAla 105
                 ::::::||||||                     ||||||
     893 CCGGCCAGGCTGAGCTCCTTCAGGCTCTCCTTGGCCCTGAGGACACGGCA 844

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro..... 120
         :::         |||   |||   :::               |||

     843 CAGATCCCCGCAGCCCTTGGCAGTGATGCCACACTCCCAGATCCACAGGG 794

     121 .....................GluHisLeuTrpGlyGlnIleValProT 130
                              |||:::   |||:::|||:::
     793 TCCTGAGCCTGGAGCTGGGGTGGAGCAGCCCTGGGCACAGCTC...CGCC 747

     130 hrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsn 146
            |||:::   |||   :::   :::::::|||   ||||||   :::
     746 ATGCCCACATCACCCAGCTTGTTGCTGCCCAGGGCCAGCTCCCGCAGCGA 697

     147 .....................ThrGlnIleArgSerValTyr........ 153
                               :::|||:::   :::|||
     696 GGCCTTGGAGGCCACAATGCCGCACAGGTCCCGGCAGTTGTCTGATGTCA 647

     154 ..........ArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla.... 165
                    :::   |||       |||   |||||||||:::
     646 CACCGCAGCTCTCCAGCTTGAGCGCCTCCAGCTGGCAGGGGGAGTCCTTC 597

     166 ..AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
           |||::::::::::|||         |||   ||||||:::   |||
     596 AGGCCCTGGCACAGCACACGGACGCCAGCCTCATTGATGTCGTTGTTGCT 547

     182 AspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCy 198
         :::   |||:::                        |||:::   ||||||
     546 AACCGTGAGCTC.....................CTTGAAGTCCGGCTTGG 518

     198 sGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                 |||||||||   :::          ||||||:::
     517 CCCT.........GAGCACGGAGGCCAG.........GGGCTCGCAGCTG 486

     215 yrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe.ArgLys...Tr 230
           :::       |||       |||:::   |||   ||| |||:::   ||
     485 GCAGCCGAGAGGCTGCAATACTCCAGCTGCAGCTTTTCCAGGCGGCACTG 436

     230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuProL 247
         ||||::::::   |||   :::::::|||   ||||||   :::   ||||
     435 GGGGTCCAGG...AGTCCTTCGCAGAGCAGCTGCAGGCCCGCATCCCCCA 389

     247 ys 247
         ||
     388 AG 387

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:AAN81083

seq_documentation_block:
ID   AAN81083 standard; DNA; 1698 BP.
XX
AC   AAN81083;
XX
DT   30-OCT-1990  (first entry)
XX
DE   Encodes human placental RNase inhibitor.
XX
KW   angiogenin inhibitory activity; neovascularisation disorders;
KW   human placental RNase inhibitor (PRI); rheumatoid arthritis;
KW   Kaposi's sarcoma; ss.
XX
OS   synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             122..1507
FT                   /*tag=  a
FT                   /product=human placental RNase inhibitor
XX
PN   EP291686-A.
XX
PD   23-NOV-1988.
XX
PF   12-APR-1988;   88EP-0105781.
XX
PR   14-APR-1987;   87US-0038008.
PR   05-APR-1988;   88US-0177942.
XX
```

```
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Shapiro R,  Vallee BL;
XX
DR   WPI; 1988-331302/47.
DR   P-PSDB; AAP80646.
XX
PT   Inhibitors of angiogenin, useful for inhibiting tumour growth -
PT   include human placental RNase inhibitor and active polypeptide
PT   segments
XX
PS   Disclosure; ; p; English.
XX
CC   cDNA sequence isolated from human placental library using probes
CC   AAN81071 to AAN81082. Tryptic digestion of the PRI protein encoded by
CC   this sequence produces peptide fragments possessing angiogenin
CC   inhibitory activity.
CC   Disorders associated with neovascularisation such as rheumatoid
CC   arthritis and Kaposi's sarcoma are treated by admin of these
CC   inhibitory peptides.
XX
SQ   Sequence 1698 BP; 316 A; 535 C; 535 G; 312 T; 0 other;


alignment_scores:
           Quality:   86.00            Length:    268
             Ratio:  0.694              Gaps:     15
 Percent Similarity:  46.269   Percent Identity: 25.373

alignment_block:
 US-09-674-779-2 x AAN81083/rev   ..

 Align seg 1/1  to reverse of: AAN81083  from: 1  to: 1698

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||:::   :::   |||   :::|||            :::::::|||::
    1222 CTGGCCCAGGCCCTGGCACAGCTCCCGCACGCCCGCATCCTCCAGCCTGT 1173

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
         :   :::        |||     |||      :::|||:::|||
    1172 TGTTGCTTATCTGTAGCTCCAGGAGAAACCTGTTCTGGGCCAGCAC.... 1127

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
                              ::::::||||||      :::
    1126 ...................TGAGCTGAAGTGGGAGCAGCAGGCGGCTGT 1097

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTy 83
         |||:::         :::|||            |||   :::::::|::
    1096 GAAGCT...GCAGGACTTCACCCACAGCGACTCCAGCTGGCAGCCAGGTT 1050

      83 rGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMet.... 98
         :|||     |||   ::::::   |||:::   :::|||   :::
    1049 CCAGCAGGGTCTCACACAGCAGTCGGGCACCCTCATCCCCCAGCTCGTTG 1000

      99 ........SerGlnLeuLeu.....................ThrThrAla 105
                :::::|||||||                      ||||||
     999 CCGGCCAGGCTGAGCTCCTTCAGGCTCTCCTTGGCCCTGAGGACACGGCA 950

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro..... 120
         :::         |||   |||   :::                 |||
     949 CAGATCCCCGCAGCCCTTGGCAGTGATGCCACACTCCCAGATCCACAGGG 900

     121 ......................GluHisLeuTrpGlyGlnIleValProT 130
                             |||:::   |||:::|||:::
     899 TCCTGAGCCTGGAGCTGGGGTGGAGCAGCCCTGGGCACAGCTC...CGCC 853

     130 hrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsn 146
             |||:::   |||   :::   :::::::|||   ||||||   :::
     852 ATGCCCACATCACCCAGCTTGTTGCTGCCCAGGGCCAGCTCCCGCAGCGA 803

     147 ....................ThrGlnIleArgSerValTyr........ 153
                              :::|||:::   :::|||
     802 GGCCTTGGAGGCCACAATGCCGCACAGGTCCCGGCAGTTGTCTGATGTCA 753

     154 ..........ArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla.... 165
                     :::   |||      |||   ||||||||||:::
     752 CACCGCAGCTCTCCAGCTTGAGCGCCTCCAGCTGGCAGGGGGAGTCCTTC 703

     166 ..AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
           |||::::::::::|||        |||   ||||||:::   |||
     702 AGGCCCTGGCACAGCACACGGACGCCAGCCTCATTGATGTCGTTGTTGCT 653

     182 AspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCy 198
         :::   |||:::                       |||:::   ||||||
     652 AACCGTGAGCTC.....................CTTGAAGTCCGGCTTGG 624

     198 sGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                 |||||||||   :::         ||||||:::
     623 CCCT.........GAGCACGGAGGCCAG.........GGGCTCGCAGCTG 592

     215 yrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe.ArgLys...Tr 230
           :::     |||      |||:::   |||   ||| |||:::   ||
     591 GCAGCCGAGAGGCTGCAATACTCCAGCTGCAGCTTTTCCAGGCGGCACTG 542

     230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuProL 247
         ||||::::::   |||   ::::::|||   ||||||   :::   ||||
     541 GGGGTCCAGG...AGTCCTTCGCAGAGCAGCTGCAGGCCCGCATCCCCCA 495

     247 ys 247
         ||
     494 AG 493

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS75095

seq_documentation_block:
ID   AAS75095 standard; cDNA; 1998 BP.
XX
AC   AAS75095;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10899.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10908.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 10899; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
```

```
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1998 BP; 311 A; 650 C; 528 G; 509 T; 0 other;


alignment_scores:
            Quality:   86.00              Length:    182
              Ratio:  1.103                Gaps:      9
 Percent Similarity:  42.857   Percent Identity: 24.725

alignment_block:
 US-09-674-779-2 x AAS75095   ..

 Align seg 1/1  to: AAS75095  from: 1  to: 1998

      87 LeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuTh 103
         |||:::|||   ||||||::::::|||   |||   :::
     177 CTCCATTCTGTACTGGGGGGCTGGCTAGGACCTGGGGCTGTGGCCTCTCA 226

     103 rThrAlaArgSer...TrpGlnAla..........................  110
           |||   |||   |||||||||
     227 GGGGGCAGCCTCTCCATGGCAGGCATCCCTGCCTTGGGCTGCCCTCCCCC 276

     111 .................CysGlyHisGluProTyrGlnLeuProProGlu 121
                          |||      |||:::|||   |||   :::
     277 AGACCCCTGACCACCCCCTGGGTCCTGTCCCCCACCAGAGCCCCAGCTCC 326

     122 HisLeuTrpGlyGlnIle......ValProThrLeuHisLeuTyrGlnAs 136
         |||||||||:::          |||   :::::::|||      :::::
     327 TGTCTGTGGGGGAGCCATCACGGTGTTCGTGCAGTCCATAGCGCTTCTCA 376

     136 pLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValT 153
         :      |||   |||      |||   :::|||||
     377 ATGTGTGTCACCCGGAACCTGCCCACAGAGAACACAG.............. 413

     153 yrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLys 169
                     :::      :::|||||:::   :::
     414 .............TGGTCTCCCCTGTCCGGGGGCGGCTTTTTCCTTCCT 449

     170 HisLeuThrAsnSerAlaIleAspIleTrp...ValProAspLeuGluIl 185
                               |||   |||||||||   :::::
     450 ............................TGGAGCGTCCCTGACGGACAAGT 472

     185 eLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpL 202
         ::::::::::::                          |||
     473 GGAGGCCTCTTGC.......................TGC.......... 488

     202 euGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGly 218
                     :::   |||   |||   ||||||   :::   |||
     489 .............GGCTGCAATGGAGGCAAGGGGCTGCAGAGCCCAGGT 524

     219 AlaIleHisLeuAspThrGlnGlyPheArg............LysTrpGl 231
         |||:::      |||   :::|||:::|||             :::|||
     525 GCACTGTGTGATGATGGGAGGGGGCTCCGTCCTGCAGGCTGGAGGTGGCA 574

     231 yAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuPro 246
         :::::::::   :::   |||:::      |||      |||
     575 TCCACACTGGACAGCTGCACTGTCCCTTCTCCTTCACTGTCCCCCA 620
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:ABL32230

seq_documentation_block:
ID   ABL32230 standard; DNA; 14568 BP.
XX
AC   ABL32230;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 203.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 203; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 14568 BP; 3168 A; 583 C; 4117 G; 6700 T; 0 other;


alignment_scores:
            Quality:   86.00              Length:    190
              Ratio:  0.915                Gaps:      9
 Percent Similarity:  49.474   Percent Identity: 23.684

alignment_block:
 US-09-674-779-2 x ABL32230/rev   ..

 Align seg 1/1  to reverse of: ABL32230  from: 1  to: 14568

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSer..... 15
         ::::::::::|||:::          ||||||   ||||||||||||
    6048 TTACGCACTTTCTCGATCGCCGCCATAACCCTCTTAATATCATCTTTAAC 5999

      16 ....MetLeuValAlaCysSerAlaProIleProThrAsnProGlnValS 31
             :::||||||:::      :::   :::         |||::::::
    5998 CCGACTCCTCGTCTCGACTCGAACCGACCTA.........CCCGACATAA 5958
```

```
     31 erProIleLys..............................ThrPro 36
          |||   |||                                 :::
   5957 TTCCGACGAAAAAACGAAACTAAAAAACGCGCCCGCGATCCTACTCGCGC 5908

     37 SerValLeuIleThrLysAspLysIleGlyAspHis..............  48
        |||:::      |||:::   :::   :::   |||
   5907 TCCGCCGACCACACTCGCGCCCGCCCGTCCGCTCACTCACTCGCGCACAA 5858

     49 .HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaH 65
         |||:::|||:::|||   :::::::::||||||:::      ::::::|
   5857 ACACTCGCACTCTCACACGCACACACTCTCACACACACACACACACACAC 5808

     65 isPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValVal 81
        ||   :::|||:::   :::            ||||||   |||::::::
   5807 ACACACACACATACACACAC............ACAAAACCCGAAACCGCC 5770

     82 ArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeu........ 95
              :::   :::::::::      :::||||||:::
   5769 GCGACCCTAACCCAAATACGCAACGCCGAAAACAACTACATCAAAACAAA 5720

     96 .............ProProMetSerGlnLeuLeuThrThr.......... 104
                     ||||||      |||:::   :::|||
   5719 AAAAACCGCGCCGCCGCCGCCGCCGCAAATCGACTCCACGAACGCGCGAC 5670

    105 ...........AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
                    |||:::      |||         :::|||   :::
   5669 CACCGAACAACCCCCGAACCTCCTACGCCCCCAAAAAATCCCCCGAACGC 5620

    118 LeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134
           |||   :::            :::   :::||||||||||:::
   5619 CCTCCTAAAAAATAACGAAAAACGCACGACACCCCGACTCTAAACGCGCG 5570

    134 r...GlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleA 150
          ::::::::::::   |||      ||||||   :::   ||||||
   5569 ATAAAAAAACGTACGACACCGCGTAACCCTTCCTCGAAAACTACAAATAC 5520

    150 rgSerValTyrArgAsnPro 156
               :::||||||
   5519 CG.........AAAAATCCA 5509
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT11658

seq_documentation_block:

```
ID   AAT11658 standard; DNA; 22481 BP.
XX
AC   AAT11658;
XX
DT   16-JAN-1997  (first entry)
XX
DE   PEDF full length sequence and flanking sequences.
XX
KW   Pigment epithelium-derived factor; PEDF; neuronal cells; neurons;
KW   glial cells; gliastatic; gliosis; central nervous system; CNS;
KW   neurodegenerative disease; injury; neuronotrophic; brain cells;
KW   Parkinson's disease; photoreceptor cells; retina; inhibition;
KW   proliferation; immunoassay; antibody; ageing; degenerative disease;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO9533480-A1.
XX
PD   14-DEC-1995.
XX
PF   06-JUN-1995;   95WO-US07201.
XX
PR   30-DEC-1994;   94US-0367841.
PR   07-JUN-1994;   94US-0257963.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Becerra SP,  Chader GJ,  Schwartz JP,  Taniwaki T;
XX
DR   WPI; 1996-039966/04.
DR   P-PSDB; AAR90287.
XX
PT   Use of pigment epithelium derived factor - for enhancing neuronal
PT   cell survival and inhibiting glial cell proliferation, useful, e.g.
PT   in CNS cell culture or to treat neuro-degenerative diseases
XX
PS   Disclosure; Page 100-122; 151pp; English.
XX
CC   Pigment epithelium-derived factor (PEDF) has both neuronotrophic and
CC   gliastatic activity, making it useful in cases where neurons die
CC   quickly and glia tend to proliferate (gliosis), e.g. in CNS cell
CC   culture, in neurodegenerative diseases and in CNS injury.  The
CC   neuronotrophic effect of PEDF is especially useful for enhancing
CC   survival of neuronal cell cultures intended for use in
CC   transplantation.  These include cultures of human foetal brain cells
CC   and neural retina and photoreceptor cells .  The gliastatic activity
CC   of PEDF can be applied to inhibiting glial cell proliferation in
CC   certain tumours.  Antibodies directed against PEDF can be used for
CC   inhibiting PEDF activity or in an immunoassay for determining
CC   levels of PEDF in fluid, cellular or tissue samples e.g for
CC   determining ageing and/or other degenerative diseases.
XX
SQ   Sequence 22481 BP; 5280 A; 5708 C; 6136 G; 5347 T; 10 other;
```

alignment_scores:

```
            Quality:   86.00              Length:    298
              Ratio:  0.662                Gaps:     14
 Percent Similarity:  43.624   Percent Identity: 21.141
```

alignment_block:

US-09-674-779-2 x AAT11658/rev ..

Align seg 1/1 to reverse of: AAT11658 from: 1 to: 22481

```
     23 ProIleProThrAsnProGlnValSerProIleLysThrProSerVal.. 38
        |||   |||   :::|||         |||   |||:::      |||
  21479 CCTGCCCCACCCCATCCA.........CCCACCAAGAGCGAAAGGGTCTT 21439

     39 .....................................LeuIleThrL 42
                                            ::::::|||:
  21438 CAGACATACTCATCTCCTGCAGGGACTTGGTGACTTCGCCTTCGTAACTC 21389

     42 ysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSer 58
        ::   ::::::||||||   :::   :::   |||      :::   |||
  21388 AGCTTCAGCTTGGGGACAGTGAGGACCGCCTGCACGGTCTTCAGTTCTCG 21339

     59 ..............HisValGlyLeuGlnAlaHisPheGluThrTrpLe 70
                     :::::::|||   :::|||   :::
  21338 GTCTATGTCATGAATGAACTCGGAGGTGAGGCTCTCCTCTATCAAGGTCA 21289

     70 uGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrL 87
           :::   |||        ||||||          :::|||
  21288 AATTCTGGGTCACTTTCAGGGGCAGGAAGAAGATGATACTCATGCTTCCG 21239

     87 euGlnSerArgLeuGlyAsnTyr.......................... 94
          |||::::::||||||||||
  21238 GTCAAGGGCAGCTGGGCAATCTGAGACAAGTGAGAAAGCATGTGGTTAGT 21189

     95 ...LeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAl 110
              |||   |||:::         ||||||||||||
  21188 CCTTCAGAGCCCAAAAGCCGGGGGACAGACACAGCCAGGAGC........ 21147

    110 aCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGln. 126
         |||       :::|||          |||            |||::::::
  21146 .TGCCTTCCCTCTCCCGTGACGGGGCCC............TGGTCTCGTC 21110

    127 IleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLe 143
        :::   |||:::   |||||||:::   :::      ||||||      ||
  21109 CTTCATCCAGCCCCTCATCTATTTTACAATTCCCCTTCCAGGTTATAGCT 21060
```

```
  143 uProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnG 160
      |   ::::::                          :::|||||||:::::::
21059 TTCTGGCAGAGAGCCA.................TCTCCTGAGGTTTCTT 21028

  160 lnCysAlaGlyGlyAlaAlaMetSer.....................Lys 169
      ::         ::::::::::::|||                     :::
21027 CAGAAAAGGATGCTGTTGGCCTTTCTTACCTTGGCCTAATACCAACTGAA 20978

  170 HisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLy 186
      :::      :::   :::   :::|||      :::||||||:::::
20977 GTTGTCTGCTGAAAGAATGTTCTCCTGTGGTTTAAGGAGCTAGAATTGAA 20928

  186 sSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCys............. 198
      :   ::::::|||   :::   :::   :::   |||
20927 TATCAGGTCTTTATTGTCTCCCAAAGTCCATCGGTGTACATTTACTGCTC 20878

  199 .............GlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPhe 210
                  :::::::|||   :::   |||:::|||:::::::::
20877 ATTTGTTAAATAGAGAGCACTGGACTAAGACAGGGCAGAACCGGGGNATA 20828

  211 GlyLeu..................................................... 212
      |||:::
20827 GGAATACCGAAGGGSCCGYAAGAGGGCTTGGAGACTTCACCCCATCCACC 20778

  213 .............GlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrG 225
                  ::::|||      |||   |||::::::|||   ::::
20777 CACCAAGAGCAAAAGCCTCCTTTGAACAGAGGCTCTTAACCTAAGGTCCC 20728

  225 lnGlyPheArgLysTrpGlyAlaGlnPheSerGluThrAsnSer...... 239
      ::|||           |||:::|||   |||:::   :::|||
20727 ATGGAACAGAT......GGTGTTCAGGGATCCCAAGAGTCCTCCGAAATC 20684

  240 ............IleCysArgHis......ValLeuProLys 247
                 :::|||   |||      |||||||||:::
20683 GTTGACAGAAACGTGTGTGCACACCTGGCGGTGCTCCCTCGG 20642
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:ABA05882

seq_documentation_block:

```
ID   ABA05882 standard; DNA; 22484 BP.
XX
AC   ABA05882;
XX
DT   21-MAR-2002  (first entry)
XX
DE   Angiogenesis inhibition method related DNA.
XX
KW   Human; antiangiogenic protein; SLED; angiogenesis; cancer; eye disorder;
KW   antitumor; antidiabetic; ophthalmological; antipsoriatic; dermatological;
KW   antiarteriosclerotic; antiarthritic; vasotropic; vulnerary;
KW   antiinflammatory; skin disease; ds.
XX
OS   Homo sapiens.
XX
PN   US2001049355-A1.
XX
PD   06-DEC-2001.
XX
PF   06-JUN-2001; 2001US-0875223.
XX
PR   23-JUL-1998;   98US-0122079.
PR   23-JUL-1997;   97US-0899304.
XX
PA   (BOUC/) BOUCK N P.
PA   (DAWS/) DAWSON D W.
PA   (GILL/) GILLIS P R.
XX
PI   Bouck NP,  Dawson DW,  Gillis PR;
XX
DR   WPI; 2002-097011/13.
XX
PT   Inhibiting angiogenesis within a tissue of eye, skin and tumor for
PT   treating diseases e.g. retinoblastoma, psoriasis and brain tumors, by
PT   providing exogenous antiangiogenic protein, SLED to the cells -
XX
PS   Disclosure; Page 8-18; 26pp; English.
XX
CC   The present invention relates to a method of inhibiting angiogenesis
CC   within a tissue by inhibiting endothelial cell migration. The method
CC   involves providing exogenous SLED (an antiangiogenic protein) to
CC   endothelial cells associated with the tissue under suitable conditions.
CC   The method is useful for inhibiting angiogenesis within a tissue
CC   preferably of eye, skin, tumour, a tissue which is within a joint, or
CC   ovarian or endometrial tissue, and especially for inhibiting the growth
CC   of a tumor. The method is particularly useful therefore for treating eye
CC   injury, hypoxia, infection, surgery, laser surgery, diabetes,
CC   retinoblastoma and preventing blindness or retarding loss of vision
CC   associated with a variety of eye diseases, disorders of skin such as
CC   psoriasis, scleroderma, tumors of skin, neovascularization as a result of
CC   infection (e.g. cat scratch disease and bacterial ulceration) or other
CC   skin disorders, tumours e.g. tumours that are difficult to operate on
CC   such as brain or prostate tumours, disorders of blood vessels such as
CC   haemangiomas and capillary proliferation within atherosclerotic plaques,
CC   muscle diseases such as myocardial angiogenesis or angiogenesis within
CC   smooth muscles, disorders of joints such as arthritis and haemophiliac
CC   joints, and other disorders associated with angiogenesis such as
CC   angiofibroma and wound granularisation.
CC   Note: The present sequence is not described further in the specification.
XX
SQ   Sequence 22484 BP; 5280 A; 5709 C; 6138 G; 5347 T; 10 other;
```

alignment_scores:

```
            Quality:   86.00              Length:    298
              Ratio:  0.662                Gaps:     14
Percent Similarity:  43.624   Percent Identity: 21.141
```

alignment_block:

US-09-674-779-2 x ABA05882/rev ..

Align seg 1/1 to reverse of: ABA05882 from: 1 to: 22484

```
   23 ProIleProThrAsnProGlnValSerProIleLysThrProSerVal.. 38
      |||   |||   :::|||         |||   |||:::      |||
21479 CCTGCCCCACCCCATCCA.........CCCACCAAGAGCGAAAGGGTCTT 21439

   39 .....................................LeuIleThrL 42
                                           ::::::|||:
21438 CAGACATACTCATCTCCTGCAGGGACTTGGTGACTTCGCCTTCGTAACTC 21389

   42 ysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSer 58
      ::   ::::::||||||   :::   :::   |||      :::   |||
21388 AGCTTCAGCTTGGGGACAGTGAGGACCGCCTGCACGGTCTTCAGTTCTCG 21339

   59 ..............HisValGlyLeuGlnAlaHisPheGluThrTrpLe 70
                     ::::::|||   :::|||   :::
21338 GTCTATGTCATGAATGAACTCGGAGGTGAGGCTCTCCTCTATCAAGGTCA 21289

   70 uGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrL 87
         :::   |||         ||||||          :::|||
21288 AATTCTGGGTCACTTTCAGGGGCAGGAAGAAGATGATACTCATGCTTCCG 21239

   87 euGlnSerArgLeuGlyAsnTyr.......................... 94
      |||::::::|||||||||
21238 GTCAAGGGCAGCTGGGCAATCTGAGACAAGTGAGAAAGCATGTGGTTAGT 21189

   95 ...LeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAl 110
             |||   |||:::         ||||||||||||
21188 CCTTCAGAGCCCAAAAGCCGGGGGACAGACACAGCCAGGAGC........ 21147

  110 aCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGln. 126
       |||      :::|||         |||            |||::::::
21146 .TGCCTTCCCTCTCCCGTGACGGGGCCC............TGGTCTCGTC 21110

  127 IleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLe 143
      :::   |||:::   ||||||:::   :::      ||||||      ||
```

```
  21109 CTTCATCCAGCCCCTCATCTATTTTACAATTCCCCTTCCAGGTTATAGCT 21060

    143 uProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnG 160
        |   ::::::                        :::||||||:::::::
  21059 TTCTGGCAGAGAGCCA.................TCTCCTGAGGTTTCTT 21028

    160 lnCysAlaGlyGlyAlaAlaMetSer.....................Lys 169
        ::        ::::::::::::|||                     :::
  21027 CAGAAAAGGATGCTGTTGGCCTTTCTTACCTTGGCCTAATACCAACTGAA 20978

    170 HisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLy 186
           :::      :::   :::   :::|||      :::|||||||:::::
  20977 GTTGTCTGCTGAAAGAATGTTCTCCTGTGGTTTAAGGAGCTAGAATTGAA 20928

    186 sSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCys............. 198
        :   :::::::|||   :::   :::   :::   |||
  20927 TATCAGGTCTTTATTGTCTCCCAAAGTCCATCGGTGTACATTTACTGCTC 20878

    199 .............GlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPhe 210
                     :::::::|||   :::   |||:::|||:::::::::
  20877 ATTTGTTAAATAGAGAGCACTGGACTAAGACAGGGCAGAACCGGGGNATA 20828

    211 GlyLeu..................................................... 212
        |||:::
  20827 GGAATACCGAAGGGSCCGYAAGAGGGCTTGGAGACTTCACCCCATCCACC 20778

    213 .............GlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrG 225
                    :::|||      |||   |||::::::|||   ::::
  20777 CACCAAGAGCAAAAGCCTCCTTTGAACAGAGGCTCTTAACCTAAGGTCCC 20728

    225 lnGlyPheArgLysTrpGlyAlaGlnPheSerGluThrAsnSer...... 239
        ::|||           |||:::|||   |||:::   :::|||
  20727 ATGGAACAGAT......GGTGTTCAGGGATCCCAAGAGTCCTCCGAAATC 20684

    240 ............IleCysArgHis......ValLeuProLys 247
                    :::|||   |||      |||||||||:::
  20683 GTTGACAGAAACGTGTGTGCACACCTGGCGGTGCTCCCTCGG 20642
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI99683

seq_documentation_block:

ID AAI99683 standard; DNA; 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 86.00 | Length: | 183 |
| Ratio: | 1.075 | Gaps: | 11 |
| Percent Similarity: | 43.716 | Percent Identity: | 24.044 |

alignment_block:
US-09-674-779-2 x AAI99683/rev ..

Align seq 1/1 to reverse of: AAI99683 from: 1 to: 4403765

```
     11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaPro...IleProTh 26
        ||||||  |||  ::::::|||::::::  ||||||  :::
4296417ACCTTGACCTCCCCGATCGTGGTCTCCGCCATAGCACCCGACGTCACCAT 4296368

     26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
        |||      ::::::|||         ||||||
4296367CAACGAGCTGATCGATCCGGCAGTGTTGCCGTCG.............. 4296334

     43 spLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHis 59
          :::|||  |||::::::  :::                 :::::|||
4296333..GAAATCCGCGATCGCTATTTGCGC.............ATCGACCAC 4296301

     60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
        |||                 ::::::|||||||||||||   |||
4296300CGCGGC.................AGCTATCTGCAGATGCATTTTGCGCT 4296269

     76 rLysGlnGlu.........ValValArgTyrGlnAlaTyr.......... 86
          |||            ::::::  |||||||||
4296268CGCGCAGCCACCGGCCTTCGCGGCGCCCTACCAGGCACTCAACGATCCGA 4296219

     87 ..LeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu 102
          :::|||:::  :::|||  :::      |||  :::
4296218GCATGCAGGCATCGATGGGCATTTTCTGTACACCGGAGCAG......... 4296178

    103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
          :::  ::::::|||:::  |||
4296177...GTTCAGCAGCAGTGGGAGGATTGC..................... 4296154

    119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136

4296154.................................................. 4296154

    136 spLeuLysSerArgGlyIleLeuProAlaAsn............ThrGln 148
                |||||||||:::||||||:::              |||
4296153........CGGCGTGGGATCGTCCCGGCCGATCCGACCGTGGTGTTGCAG 4296112

    149 IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAl 165
        |||  |||::::::  :::|||:::|||      ::::::    ||
4296111ATCCCGTCGCTGCAT...GATCCCAGCCTGGCCCCGGCGGGTAAGCAGGC 4296065

    165 aAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
        ||||                    ||||||:::  :::|||  |||
4296064CGCG.................TCGGCGTTCGCAATGTGGTTCCCG 4296037
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI99682

seq_documentation_block:
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;


alignment_scores:
            Quality:   86.00              Length:    183
              Ratio:  1.075                Gaps:     11
 Percent Similarity:  43.716   Percent Identity: 24.044

alignment_block:
 US-09-674-779-2 x AAI99682/rev   ..

 Align seg 1/1  to reverse of: AAI99682  from: 1  to: 4411529

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaPro...IleProTh 26
         ||||||   |||   ::::::|||::::::   ||||||   :::
  4304162ACCTTGACCTCCCCGATCGTGGTCTCCGCCATAGCACCCGACGTCACCAT 4304113

      26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
         |||      ::::::|||         ||||||
  4304112CAACGAGCTGATCGATCCGGCAGTGTTGCCGTCG.............. 4304079

      43 spLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHis 59
            :::|||   |||::::::   :::                 ::::::|||
  4304078..GAAATCCGCGATCGCTATTTGCGC.............ATCGACCAC 4304046

      60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
         |||                  ::::::||||||||||||   |||
  4304045CGCGGC.................AGCTATCTGCAGATGCATTTTGCGCT 4304014

      76 rLysGlnGlu.........ValValArgTyrGlnAlaTyr.......... 86
         |||            ::::::   |||||||||
  4304013CGCGCAGCCACCGGCCTTCGCGGCGCCCTACCAGGCACTCAACGATCCGA 4303964

      87 ..LeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu 102
           :::|||:::   :::|||   :::.      |||   :::
  4303963GCATGCAGGCATCGATGGGCATTTTCTGTACACCGGAGCAG......... 4303923

     103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
         :::   ::::::|||:::   |||
  4303922...GTTCAGCAGCAGTGGGAGGATTGC....................... 4303899

     119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136

  4303899.................................................. 4303899

     136 spLeuLysSerArgGlyIleLeuProAlaAsn.............ThrGln 148
                 |||||||||:::||||||:::                |||
  4303898........CGGCGTGGGATCGTCCCGGCCGATCCGACCGTGGTGTTGCAG 4303857

     149 IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAl 165
         |||   |||::::::   :::|||:::|||       ::::::      ||
  4303856ATCCCGTCGCTGCAT...GATCCCAGCCTGGCCCCGGCGGGTAAGCAGGC 4303810

     165 aAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
         ||||                    ||||||:::   :::|||   |||
  4303809CGCG.....................TCGGCGTTCGCAATGTGGTTCCCG 4303782

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAZ53602

seq_documentation_block:
ID   AAZ53602 standard; DNA; 915 BP.
XX
AC   AAZ53602;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 280 partial DNA sequence SEQ ID NO:1153.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
```

```
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY74840.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 7; Page 636; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 915 BP; 277 A; 292 C; 195 G; 151 T; 0 other;


alignment_scores:
            Quality:   85.50              Length:    215
              Ratio:  0.864                Gaps:     12
 Percent Similarity:  46.047   Percent Identity: 22.791

alignment_block:
 US-09-674-779-2 x AAZ53602   ..

 Align seg 1/1  to: AAZ53602  from: 1  to: 915

      17 LeuValAlaCysSerAla.....................ProIleProTh 26
         |||      |||:::|||                     |||||||||
     248 CTGCCGATGTGCAACGCGCCGTCAAACAAAGCAAAGTATCCTATACCGAA 297

      26 rAsnProGlnValSer..ProIleLysThrProSerValLeuIleThrLy 42
          :::|||:::::::|||  |||:::|||                 :::::
     298 GCGACCAAAGGCATCCAACCCCTCAAA..................GCCGA 329

      42 sAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerH 59
         :::::::   |||   ||||||   |||:::||||||::::::      |
     330 AGAAGAAGGCGGACACCATCACGACCACGATCATGACCACGAAGGACACC 379

      59 isValGlyLeuGlnAlaHisPheGlu............ThrTrpLeuGln 71
         ||         :::   |||   |||            :::|||   :::
     380 AC.........CATGACCACGGCGAATATGACCCGCACGTCTGGAACGAC 420

      72 ......MetHisHisAlaThrLysGlnGluValValArgTyr........ 83
               :::      |||      |||:::|||::::::
     421 CCCGTCCTTATGTCCGCCTATGCCCAAAACGTTGCCAAAGCCCTGATAAA 470

      84 ................GlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrL 95
                         ::::::|||   |||:::||||||||||||||||
     471 GGCCGATCCCGAAGGCAAAGTTTATTATCAACAACGCTTGGGCAACTACC 520

      95 euProProMetSerGlnLeu............................. 101
                ::::::::::|||
     521 AAATGCAGCTCAAAAAACTGCACAGCGACGCACAAGCCGCATTTAATGCC 570

     102 LeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln.. 117
         :::   :::|||:::               |||||||:::   ::::::
     571 GTCCCTGCTGCCAAACGCAAAGTCCTGACCGGGCACGATGCCTTTTCCTA 620

     118 ............................LeuProProGluHisLeuTrpG 125
                                     :::   |||:::   :::   :
     621 TATGGGCAAACGTTACCATATCGAATTCATCGCCCCGCAAGGCGTGAGCA 670

     125 lyGlnIleValProThrLeuHis............LeuTyrGlnAspLeu 137

         :::::      |||:::                    :::   :::::::::
     671 GCGAAGCCGAGCCTTCGGCCAAACAAGTCGCCGCCATCATCCGACAAATC 720

     138 LysSerArgGlyIleLeuProAlaAsnThrGln.............IleAr 150
         |||   :::|||||||      :::   |||:::               ||
     721 AAACGCGAAGGCATCAAAGCCGTCTTTACCGAAAACATCAAGGACACCCG 770

     150 gSerValTyrArg...AsnProGluLeuAsnGlnCysAlaGlyGlyAlaA 166
         |   |||   |||          |||   :::      ::::::|||
     771 TATGGTTGACCGTATCGCCAAAGAAACCGGTGTCAACGTCAGCGGCAAAC 820

     166 laMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpVal 180
            |||      |||   |||:::      |||   ::::::
     821 TGTATTCCGACGCACTCGGCAACGCGCCCGCAGACACCTACATC 864

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:AAQ46887

seq_documentation_block:
ID   AAQ46887 standard; DNA; 2188 BP.
XX
AC   AAQ46887;
XX
DT   08-FEB-1994  (first entry)
XX
DE   Sequence of chicken brain cDNA encoding aceylcholine receptor-inducing
DE   activity (ARIA) protein E18.
XX
KW   acetylcholine receptor-inducing activity protein; ARIA; E18; ss.
XX
OS   Gallus.
XX
FH   Key             Location/Qualifiers
FT   CDS             172..972
FT                   /*tag=  a
XX
PN   US5237056-A.
XX
PD   17-AUG-1993.
XX
PF   29-MAY-1991;   91US-0706872.
XX
PR   29-MAY-1991;   91US-0706872.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Fischbach GD;
XX
DR   WPI; 1993-272184/34.
DR   P-PSDB; AAR40171.
XX
PT   DNA encoding protein which co-purifies with acetyl:choline
PT   receptor-inducing activity - used to obtain prods. for treating
PT   deficiencies in neural transmitter receptors, e.g. Alzheimer's
PT   disease
XX
PS   Claim 2; Figure 2; 17pp; English.
XX
CC   ARIA promotes the synthesis and accumulation of acetylcholine
CC   receptors (AChRs) in muscle cells. The protein was isolated on the
CC   basis of its ability to increase the rate of insertion of AChRs into
CC   the surface membrane of chick myotubes. Oligos with sequences
CC   corresp. to the sequence of the protein were used to amplify from
CC   chicken brain cDNA a 34-nucleotide sequence encoding 11 AAs from
CC   ARIA. The 34-nucleotide sequence corresp. to the 3' nucleotides of
CC   the 7th AA codon through the 3' nucleotide of the 18th AA codon.
CC   Each primer contd. an EcoRI site at its 5' end. The first pair of
CC   oligos are AAQ46882/Q46883. The second pair are AAQ46884/Q46885. The
CC   oligos are nested pairs. The PCR amplified prod. was used as a
CC   probe to screen an E18 chick brain cDNA library. AAQ46887 is the
CC   insert from a positive clone obtd. by the screening an E18 chick
CC   brain cDNA library in lambda gt10 and introduced into a
CC   plasmid designated p65-21.
XX
```

SQ Sequence 2188 BP; 496 A; 656 C; 572 G; 464 T; 0 other;

alignment_scores:
Quality: 85.50 Length: 207
Ratio: 0.830 Gaps: 12
Percent Similarity: 49.758 Percent Identity: 24.638

alignment_block:
US-09-674-779-2 x AAQ46887 ..

Align seg 1/1 to: AAQ46887 from: 1 to: 2188

```
  21 SerAlaProIleProThrAsnProGlnValSerProIleLysThrProSe 37
     |||:::|||:::|||   :::     |||:::          ::
 691 AGCAGCCCCGTGCCACAGGACGTCTTCGTGGCCGATTGCTTT......AA 734

  37 rValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisA 54
     ::::   :::|||:::   :::|||||||                :::   :
 735 CATCACAGTGACTGAGTACAGCATTGGCCCTGCTGCCAAGAAGAACACCT 784

  54 spGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeu 70
     ::|||:::|||:::              |||      :::|||   :::
 785 CCGAGGCTGTGGCG..............GCAGCAAACCAAACGGAGGTG 819

  71 GlnMetHisHisAlaThrLysGlnGluValValArg.........TyrGl 84
     :::|||::::::   :::      :::|||:::|||           ||
 820 GAGATGGAGAACAAAGTGGTGACGAAGGTGATCCGCGAGATGTGCGTGCA 869

  84 nAlaTyrLeuGlnSerArgLeuGlyAsn..................... 93
     |   |||   :::   ||||||::::::
 870 GCAGTACCGCGAGTACCGCCTGGCCTCGGGCATCCAGCTGCACCCTGCTG 919

  94 .....TyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSer... 107
          :::|||      :::      ||||||||||||     :::
 920 ACACCTGGCTCGCCGTCCTCCTCCTCCTCCTCACCACCCTTTTTGCCATG 969

 108 ......TrpGlnAla............CysGlyHisGluProTyrGlnLe 118
           |||:::|||            ||||||   |||      :::
 970 CACTGATGGGATGCCGTGCCCCGGCCCTGTGGCAGTGAGATGACATCGTG 1019

 118 uProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrG 135
      |||      |||   ||||||      ||||||      ::::::
1020 TCCCCGTGCCCACCCATGGGGT......GTTCCTTGTCCTCGCTTTTGTC 1063

 135 lnAspLeu......LysSerArgGlyIleLeuProAlaAsnThrGlnIle 149
       :::|||         |||         |||||||||:::::::::::::
1064 CATCTTTGGTGAAGATGTCCCCCCGCTGCCTCCCCGCAGGCTCTGATTTG 1113

 150 .........ArgSerValTyrArgAsnProGluLeuAsnGlnCysAla.. 162
             :::::::::         |||      ::::::   |||
1114 GGGAAATGGGAGGGGATTTTGTCCTGTCCTGGTCGTGGCAGGACGGCTGC 1163

 163 ....GlyGlyAlaAlaMetSerLysHis.....................L 171
         ||||||::::::      |||:::                     |
1164 TGGTGGTGGAGTGGGATGCCCAAAAAATGGCCTTCACCACTTCCTCCTCC 1213

 171 euThrAsnSerAlaIleAspIleTrpVal...ProAspLeuGluIleLys 186
     ||      |||:::   :::||||||:::   |||   |||   :::
1214 TCTTCCTTTCTGGGGCGGAGATATGGGCTCGTCCAGCCCTTATTGTCCCT 1263

 187 SerGlnAlaLeuTyrGluLeu 193
     ::::::|||      ::::::
1264 GCAAGAGCGTATCTGAAAATC 1284
```

OM of: US-09-674-779-2 to: Issued_Patents_NA:* out_format : pfs

Date: Sep 16, 2002 6:53 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124121_25853/app_query.fasta_1.310
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=std.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_47 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 45.440000

score_list:

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|---|---|---|---|---|---|---|
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-801-263A-4 | + | 92.00 | 132.77 | 10.46 | 11717 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-102-248-4 | + | 92.00 | 132.77 | 10.46 | 11717 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-910-731-5 | - | 86.00 | 146.30 | 1.85 | 1386 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-367-841A-43 | - | 86.00 | 112.24 | 145.63 | 22481 | |
| /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-07201-43 | - | 86.00 | 112.24 | 145.63 | 2248 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2 | - | 86.00 | 47.73 | 4.0e+05 | 4403765 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1 | - | 86.00 | 47.71 | 4.0e+05 | 4411529 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-706-872-2 | + | 85.50 | 139.67 | 4.32 | 2188 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-673-312-9 | - | 85.00 | 125.96 | 25.04 | 6162 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-276-099A-1 | - | 83.50 | 131.43 | 12.42 | 3046 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-781-890-1 | - | 83.50 | 131.43 | 12.42 | 3046 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-087-465-11 | - | 83.50 | 131.43 | 12.42 | 3046 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-621-193A-1 | + | 82.50 | 124.50 | 30.23 | 4527 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-018-489C-1 | + | 82.50 | 124.50 | 30.23 | 4527 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-23 | + | 82.00 | 144.08 | 2.45 | 837 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-36 | + | 82.00 | 142.83 | 2.88 | 927 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-103 | + | 82.00 | 142.83 | 2.88 | 927 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-31 | + | 82.00 | 135.37 | 7.49 | 1707 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-038-832-3 | + | 81.50 | 135.14 | 7.72 | 1597 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-038-832-1 | + | 81.50 | 127.85 | 19.67 | 2900 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-439-313-474 | + | 81.00 | 134.11 | 8.81 | 1594 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-741-881-1 | + | 81.00 | 105.43 | 348.62 | 16656 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-739-158-1 | + | 81.00 | 105.43 | 348.62 | 16656 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-739-167-1 | + | 81.00 | 105.43 | 348.62 | 16656 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-404-796-1 | + | 81.00 | 105.43 | 348.62 | 16656 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-931-869-1 | + | 81.00 | 105.43 | 348.62 | 16656 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-350-399-1 | + | 81.00 | 105.43 | 348.62 | 16656 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-276-452A-66 | + | 80.50 | 140.15 | 4.06 | 893 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-798-744-66 | + | 80.50 | 140.15 | 4.06 | 893 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-203-532F-3 | + | 80.50 | 139.51 | 4.41 | 941 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-078-465-3 | + | 80.50 | 139.51 | 4.41 | 941 | |
| /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-01882A-3 | + | 80.50 | 139.51 | 4.41 | 941 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-524-2 | + | 80.00 | 138.48 | 5.03 | 940 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-524-8 | - | 80.00 | 138.48 | 5.03 | 940 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-846-762-1 | - | 80.00 | 98.66 | 830.70 | 24417 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-440-845D-8 | - | 79.50 | 133.46 | 9.57 | 1300 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-868-458-8 | - | 79.50 | 133.46 | 9.57 | 1300 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-801-263A-8 | + | 79.50 | 106.60 | 299.96 | 11703 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-102-248-8 | + | 79.50 | 106.60 | 299.96 | 11703 | |
| /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-00362-4 | + | 79.00 | 133.69 | 9.29 | 1171 | |
| /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-165-241-2 | + | 78.00 | 127.42 | 20.77 | 1648 | |
| /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-120-1 | + | 78.00 | 125.80 | 25.57 | 1882 | |
| /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-867-970-1 | + | 78.00 | 125.80 | 25.57 | 1882 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-326-217-1 | + | 78.00 | 125.80 | 25.57 | 1882 | |
| /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-972-218-1 | + | 77.50 | 100.31 | 672.60 | 13905 | |

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-801-263A-4

seq_documentation_block:
; Sequence 4, Application US/08801263A
; Patent No. 5811407
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5811407th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,263A
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11717 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-801-263A-4

alignment_scores:
Quality: 92.00 Length: 283
Ratio: 0.724 Gaps: 18
Percent Similarity: 44.876 Percent Identity: 24.735

alignment_block:
US-09-674-779-2 x US-08-801-263A-4 ..

Align seg 1/1 to: US-08-801-263A-4 from: 1 to: 11717

```
     11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
        ||||||   |||      |||   :::||||||            |||::
   8664 ACCTTGACCAGCCCGTACTTGGGCACATGCTCGTACTGTCACCATACTGA 8713

     27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
        :|||      ||||||||||||      :::|||            |||:
   8714 ACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGG.....GATG 8754

     44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
        ::      |||  ||||||||||||        :::   |||
   8755 AAGCGGACGACAACACCATACGCATACAGACTTCCGCCCAGTT..... 8797
```

```
    60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                              |||::::::      :::::::::
  8798 .........................TGGATACGACCAAAGCGGAGCAGC 8821

    77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln............. 88
       |||   :::   ||| ||||||::::::      |||
  8822 AAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACCGTCA 8871

    88 ................................................. 88

  8872 AAGAAGGCACTATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGA 8921

    89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
          |||      |||   :::|||           ||||||
  8922 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGA 8971

    98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
         ::::::      ::::::::::::::::|||      :::|||
  8972 CAGCGTAACGGTTAGTATAGCGAGTAGCAACTCAGCAACGTCATGCACAA 9021

   112 ............................GlyHisGluProTyrGlnLeu 118
                                  |||:::|||   |||:::|||
  9022 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGACCTA 9071

   119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
       ||||||   |||      |||:::   :::|||         :::|||::
  9072 CCTCCCGTTCAC......GGTAAGAAGATTCCTTGCACA...GTGTACGA 9112

   135 nAspLeuLys......SerArgGlyIleLeu............ProAla. 145
       :   ||||||      :::   |||   :::                |||:::
  9113 CCGTCTGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGAC 9162

   146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
       :::|||   ||||||   ::::::|||||||   :::
  9163 CGCACGCCTATACGTCCTATCTGGAGGAATCATCAG.............. 9198

   162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
        |||   ::::::   |||   |||
  9199 .GGAAAGTCTACGCGAAGCCACCAT........................ 9222

   179 rpValProAspLeuGluIleLysSerGlnAla.LeuTyrGluLeuGlnAs 195
          |||:::      :::::::::::::||| :::   :::|||||||::
  9223 .....CCGGAAAGAACATTACGTACGAGTGCAAGTGCGGCGATTACAAGA 9267

   195 nArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212
       :|||      |||      :::      :::|||:::
  9268 CCGG.........TACCGTTACGACCCGTACCGAAATCAC.......... 9298

   212 euGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg 228
         ||||||:::        |||::::::         :::      :::
  9299 ..GGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAAC 9346

   229 LysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArg 242
       :::   |||   ||||||:::      ::::::   ||||||
  9347 GAAGTGGGTCTTCAATTCGCCGGACTTGATCAGACATGCCGA 9388
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-102-248-4

```
seq_documentation_block:
; Sequence 4, Application US/09102248
; Patent No. 6008035
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6008035th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,248
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11717 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-102-248-4


alignment_scores:
            Quality:   92.00              Length:    283
              Ratio:  0.724                Gaps:     18
 Percent Similarity:  44.876   Percent Identity: 24.735

alignment_block:
 US-09-674-779-2 x US-09-102-248-4    ..

 Align seg 1/1  to: US-09-102-248-4  from: 1  to: 11717

     11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
        ||||||   |||      |||   :::|||||||              |||::
   8664 ACCTTGACCAGCCCGTACTTGGGCACATGCTCGTACTGTCACCATACTGA 8713

     27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
        :|||      ||||||||||||      :::|||              |||:
   8714 ACCGTGCTTTAGCCCGATTAAGATCGAGCAGGTCTGG.........GATG 8754

     44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
        ::      |||  ||||||||||||          :::   |||
   8755 AAGCGGACGACAACACCATACGCATACAGACTTCCGCCCAGTT....... 8797

     60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                 |||::::::      :::::::::
   8798 .........................TGGATACGACCAAAGCGGAGCAGC 8821

     77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln............. 88
        |||   :::   ||| ||||||::::::      |||
   8822 AAGCTCAAATAAGTACCGCTACATGTCGCTCGAGCAGGATCATACCGTCA 8871

     88 ................................................. 88

   8872 AAGAAGGCACTATGGATGACATCAAGATCAGCACCTCAGGACCGTGTAGA 8921

     89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
           |||      |||   :::|||           ||||||
   8922 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCGAAGTGTCCTCCAGGGGA 8971

     98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
          ::::::      ::::::::::::::::|||      :::|||
   8972 CAGCGTAACGGTTAGTATAGCGAGTAGCAACTCAGCAACGTCATGCACAA 9021
```

```
     112 ..........................GlyHisGluProTyrGlnLeu 118
                                    |||:::|||   |||:::|||
    9022 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGACCTA 9071

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
         ||||||   |||      |||:::   :::|||         :::||||::
    9072 CCTCCCGTTCAC......GGTAAGAAGATTCCTTGCACA...GTGTACGA 9112

     135 nAspLeuLys......SerArgGlyIleLeu............ProAla. 145
         :   ||||||       :::   |||   :::            |||:::
    9113 CCGTCTGAAAGAAACAACCGCCGGCTACATCACTATGCACAGGCCGGGAC 9162

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
         :::|||   ||||||   ::::::||||||   :::
    9163 CGCACGCCTATACGTCCTATCTGGAGGAATCATCAG.............. 9198

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
          |||   ::::::   |||   |||
    9199 .GGAAAGTCTACGCGAAGCCACCAT........................ 9222

     179 rpValProAspLeuGluIleLysSerGlnAla.LeuTyrGluLeuGlnAs 195
           |||:::      ::::::::::::|||  :::   :::||||||::
    9223 .....CCGGAAAGAACATTACGTACGAGTGCAAGTGCGGCGATTACAAGA 9267

     195 nArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212
         :|||         |||      :::      :::|||:::
    9268 CCGG.........TACCGTTACGACCCGTACCGAAATCAC.......... 9298

     212 euGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg 228
           |||||||:::         |||::::::         :::      :::
    9299 ..GGGCTGCACCGCCATCAAGCAGTGCGTCGCCTATAAGAGCGACCAAAC 9346

     229 LysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArg 242
         :::   |||   ||||||:::      ::::::   ||||||
    9347 GAAGTGGGTCTTCAATTCGCCGGACTTGATCAGACATGCCGA 9388

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-910-731-5

seq_documentation_block:
; Sequence 5, Application US/08910731
; Patent No. 5932440
;  GENERAL INFORMATION:
;    APPLICANT:  CHATTERJEE, DEB K.
;    APPLICANT:  SHANDILYA, HARINI
;    TITLE OF INVENTION:  Mammalian Ribonuclease Inhibitors and Use Thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVE., N.W., SUITE 600
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,731
;      FILING DATE:  (Herewith)
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/795,395
;      FILING DATE:  04-FEB-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/794,546
;      FILING DATE:  03-FEB-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/024,057
;      FILING DATE:  16-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ESMOND, ROBERT W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0942.3440003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
US-08-910-731-5


alignment_scores:
            Quality:   86.00              Length:    268
              Ratio:  0.694                Gaps:     15
 Percent Similarity:  46.269   Percent Identity: 25.373

alignment_block:
 US-09-674-779-2 x US-08-910-731-5/rev   ..

 Align seg 1/1  to reverse of: US-08-910-731-5  from: 1  to: 1386

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||:::   :::   |||   :::|||           ::::::|||::
    1101 CTGGCCCAGGCCCTGGCACAGCTCCCGCACGCCCGCATCCTCCAGCCTGT 1052

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisT 50
         :   :::         |||      |||      :::|||:::|||
    1051 TGTTGCTTATCTGTAGCTCCAGGAGAAACCTGTTCTGGGCCAGCAC.... 1006

      50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
                                  ::::::||||||      :::
    1005 ...................TGAGCTGAAGTGGGGGCAGCAGGCGGCTGT 976

      67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTy 83
         |||:::          :::|||            |||   :::::|||::
     975 GAAGCT...GCAGGACTTCACCCACAGCGACTCCAGCTGGCAGCCAGGTT 929

      83 rGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMet.... 98
         :|||      |||   ::::::   |||:::   :::|||   :::
     928 CCAGCAGGGTCTCACACAGCAGTCGGGCACCCTCATCCCCCAGCTCGTTG 879

      99 ........SerGlnLeuLeu.....................ThrThrAla 105
                 ::::::||||||                      ||||||
     878 CCGGCCAGGCTGAGCTCCTTCAGGCTCTCCTTGGCCCTGAGGACACGGCA 829

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro..... 120
         :::          |||   |||   :::            |||
     828 CAGATCCCCGCAGCCCTTGGCAGTGATGCCACACTCCCAGATCCACAGGG 779

     121 ......................GluHisLeuTrpGlyGlnIleValProT 130
                               |||:::   |||:::|||:::
     778 TCCTGAGCCTGGAGCTGGGGTGGAGCAGCCCTGGGCACAGCTC...CGCC 732

     130 hrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsn 146
           |||:::   |||   :::   :::::::|||   ||||||   :::
     731 ATGCCCACATCACCCAGCTTGTTGCTGCCCAGGGCCAGCTCCCGCAGCGA 682

     147 ......................ThrGlnIleArgSerValTyr........ 153
                                 :::|||:::   :::|||
     681 GGCCTTGGAGGCCACAATGCCGCACAGGTCCCGGCAGTTGTCTGATGTCA 632

     154 ..........ArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla.... 165
                     :::   |||      |||   ||||||||||:::
     631 CACCGCAGCTCTCCAGCTTGAGCGCCTCCAGCTGGCAGGGGGAGTCCTTC 582

     166 ..AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
           |||::::::::::|||         |||   ||||||:::   |||
     581 AGGCCCTGGCACAGCACACGGACGCCAGCCTCATTGATGTCGTTGTTGCT 532
```

```
    182 AspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCy 198
        :::   |||:::                  |||:::   ||||||
    531 AACCGTGAGCTC...................CTTGAAGTCCGGCTTGG 503

    198 sGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
                  |||||||||   :::         |||||||:::
    502 CCCT.........GAGCACGGAGGCCAG.........GGGCTCGCAGCTG 471

    215 yrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe.ArgLys...Tr 230
          :::      |||      |||:::   |||   ||| |||:::   ||
    470 GCAGCCGAGAGGCTGCAATACTCCAGCTGCAGCTTTTCCAGGCGGCACTG 421

    230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuProL 247
        ||||::::::   |||   ::::::|||   ||||||   :::   ||||
    420 GGGGTCCAGG...AGTCCTTCGCAGAGCAGCTGCAGGCCCGCATCCCCCA 374

    247 ys 247
        ||
    373 AG 372

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-367-841A-43

seq_documentation_block:
; Sequence 43, Application US/08367841A
; Patent No. 6319687
;  GENERAL INFORMATION:
;    APPLICANT:  Chader, Gerald J.; Rodriguez,
;    APPLICANT:  Ignacio R.; Mazuruk, Krzysztof;
;    APPLICANT:  Tombran-Tink, Joyce
;    TITLE OF INVENTION:  PIGMENT EPITHELIUM
;    TITLE OF INVENTION:  DERIVED FACTOR: CHARACTERIZATION GENOMIC
;    TITLE OF INVENTION:  ORGANIZATION AND SEQUENCE OF THE PEDF GENE
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morgan & Finnegan
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/367,841A
;      FILING DATE:  30-DEC-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/257,963
;      FILING DATE:  07-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/952,796
;      FILING DATE:  24-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DOROTHY R. AUTH
;      REGISTRATION NUMBER:  36434
;      REFERENCE/DOCKET NUMBER:  20264126US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22481 Base Pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  P1-147
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  full length genomic
;      OTHER INFORMATION:  sequence for PEDF plus flanking sequences.
US-08-367-841A-43


alignment_scores:
            Quality:   86.00              Length:    298
              Ratio:  0.662                Gaps:     14
 Percent Similarity:  43.624   Percent Identity: 21.141

alignment_block:
 US-09-674-779-2 x US-08-367-841A-43/rev   ..

 Align seg 1/1  to reverse of: US-08-367-841A-43  from: 1  to: 22481

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSerVal.. 38
         |||   |||   :::|||         |||   |||:::      |||
  21479 CCTGCCCCACCCCATCCA.........CCCACCAAGAGCGAAAGGGTCTT 21439

      39 .....................................LeuIleThrL 42
                                              :::::::|||:
  21438 CAGACATACTCATCTCCTGCAGGGACTTGGTGACTTCGCCTTCGTAACTC 21389

      42 ysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSer 58
         ::   ::::::||||||   :::   :::   |||      :::   |||
  21388 AGCTTCAGCTTGGGGACAGTGAGGACCGCCTGCACGGTCTTCAGTTCTCG 21339

      59 ..............HisValGlyLeuGlnAlaHisPheGluThrTrpLe 70
                       :::::::|||   :::|||   :::
  21338 GTCTATGTCATGAATGAACTCGGAGGTGAGGCTCTCCTCTATCAAGGTCA 21289

      70 uGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrL 87
           :::   |||         ||||||              :::|||
  21288 AATTCTGGGTCACTTTCAGGGGCAGGAAGAAGATGATACTCATGCTTCCG 21239

      87 euGlnSerArgLeuGlyAsnTyr........................... 94
          |||::::::|||||||||
  21238 GTCAAGGGCAGCTGGGCAATCTGAGACAAGTGAGAAAGCATGTGGTTAGT 21189

      95 ...LeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAl 110
               |||   |||:::         ||||||||||||
  21188 CCTTCAGAGCCCAAAAGCCGGGGGACAGACACAGCCAGGAGC........ 21147

     110 aCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGln. 126
          |||       :::|||         |||            |||::::::
  21146 .TGCCTTCCCTCTCCCGTGACGGGGCCC............TGGTCTCGTC 21110

     127 IleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLe 143
         :::   |||:::   |||||||:::   :::      ||||||      ||
  21109 CTTCATCCAGCCCCTCATCTATTTTACAATTCCCCTTCCAGGTTATAGCT 21060

     143 uProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnG 160
         |   ::::::                        :::||||||:::::::
  21059 TTCTGGCAGAGAGCCA.................TCTCCTGAGGTTTCTT 21028

     160 lnCysAlaGlyGlyAlaAlaMetSer.....................Lys 169
         ::          ::::::::::::|||                     :::
  21027 CAGAAAAGGATGCTGTTGGCCTTTCTTACCTTGGCCTAATACCAACTGAA 20978

     170 HisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLy 186
            :::      :::   :::   :::|||      :::||||||:::::
  20977 GTTGTCTGCTGAAAGAATGTTCTCCTGTGGTTTAAGGAGCTAGAATTGAA 20928

     186 sSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCys.............. 198
         :   ::::::|||   :::   :::   :::   |||
  20927 TATCAGGTCTTTATTGTCTCCCAAAGTCCATCGGTGTACATTTACTGCTC 20878

     199 ..............GlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPhe 210
                       :::::::|||   :::   |||:::|||:::::::::
  20877 ATTTGTTAAATAGAGAGCACTGGACTAAGACAGGGCAGAACCGGGGNATA 20828

     211 GlyLeu.......................................... 212
         |||:::
```

```
    20827 GGAATACCGAAGGGSCCGYAAGAGGGCTTGGAGACTTCACCCCATCCACC 20778

      213 .............GlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrG 225
                      :::|||      |||   |||::::::|||   ::::
    20777 CACCAAGAGCAAAAGCCTCCTTTGAACAGAGGCTCTTAACCTAAGGTCCC 20728

      225 lnGlyPheArgLysTrpGlyAlaGlnPheSerGluThrAsnSer..... 239
          ::|||           |||:::|||   |||:::   :::|||
    20727 ATGGAACAGAT......GGTGTTCAGGGATCCCAAGAGTCCTCCGAAATC 20684

      240 ............IleCysArgHis......ValLeuProLys 247
                    :::|||   |||      |||||||||:::
    20683 GTTGACAGAAACGTGTGTGCACACCTGGCGGTGCTCCCTCGG 20642

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-07201-43

seq_documentation_block:
; Sequence 43, Application PC/TUS9507201
;  GENERAL INFORMATION:
;     APPLICANT:  Chader, Gerald J.; Becerra, Sofia
;     APPLICANT:  Patricia; Schwartz, Joan P.;
;     APPLICANT:  Taniwaki, Takayuki
;     TITLE OF INVENTION:  PIGMENT EPITHELIUM
;     TITLE OF INVENTION:  DERIVED FACTOR: CHARACTERIZATION GENOMIC
;     TITLE OF INVENTION:  ORGANIZATION AND SEQUENCE OF THE PEDF GENE
;     NUMBER OF SEQUENCES:  43
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Morgan & Finnegan, L.L.P.
;       STREET:  345 Park Avenue
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy Disk
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORDPERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/07201
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/367,841
;       FILING DATE:  30-DEC-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/257,963
;       FILING DATE:  07-JUN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/952,796
;       FILING DATE:  24-SEP-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  DOROTHY R. AUTH
;       REGISTRATION NUMBER:  36434
;       REFERENCE/DOCKET NUMBER:  20264126PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;   INFORMATION FOR SEQ ID NO:  43:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  22481 Base Pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Double
;       TOPOLOGY:  Unknown
;     MOLECULE TYPE:  Genomic DNA
;     FEATURE:
;       NAME/KEY:  Pl-147
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:  full length genomic
;       OTHER INFORMATION:  sequence for PEDF plus flanking sequences.
PCT-US95-07201-43

alignment_scores:
           Quality:   86.00              Length:    298
             Ratio:  0.662                Gaps:     14
 Percent Similarity:  43.624   Percent Identity: 21.141

alignment_block:
 US-09-674-779-2 x PCT-US95-07201-43/rev   ..

 Align seg 1/1  to reverse of: PCT-US95-07201-43  from: 1  to: 22481

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSerVal.. 38
         |||   |||   :::|||         |||   |||:::      |||
  21479 CCTGCCCCACCCCATCCA.........CCCACCAAGAGCGAAAGGGTCTT 21439

      39 .....................................LeuIleThrL 42
                                             ::::::|||:
  21438 CAGACATACTCATCTCCTGCAGGGACTTGGTGACTTCGCCTTCGTAACTC 21389

      42 ysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSer 58
         ::   ::::::||||||   :::   :::   |||      :::   |||
  21388 AGCTTCAGCTTGGGGACAGTGAGGACCGCCTGCACGGTCTTCAGTTCTCG 21339

      59 ..............HisValGlyLeuGlnAlaHisPheGluThrTrpLe 70
                       ::::::|||   :::|||   :::
  21338 GTCTATGTCATGAATGAACTCGGAGGTGAGGCTCTCCTCTATCAAGGTCA 21289

      70 uGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrL 87
           :::   |||         ||||||              :::|||
  21288 AATTCTGGGTCACTTTCAGGGGCAGGAAGAAGATGATACTCATGCTTCCG 21239

      87 euGlnSerArgLeuGlyAsnTyr........................... 94
         |||::::::|||||||||
  21238 GTCAAGGGCAGCTGGGCAATCTGAGACAAGTGAGAAAGCATGTGGTTAGT 21189

      95 ...LeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAl 110
               |||   |||:::         ||||||||||||
  21188 CCTTCAGAGCCCAAAAGCCGGGGGACAGACACAGCCAGGAGC........ 21147

     110 aCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGln. 126
          |||      :::|||         |||            |||::::::
  21146 .TGCCTTCCCTCTCCCGTGACGGGGCCC.............TGGTCTCGTC 21110

     127 IleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLe 143
         :::   |||:::   ||||||:::   :::      ||||||     ||
  21109 CTTCATCCAGCCCCTCATCTATTTTACAATTCCCCTTCCAGGTTATAGCT 21060

     143 uProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnG 160
         |   ::::::                        :::||||||:::::::
  21059 TTCTGGCAGAGAGCCA.................TCTCCTGAGGTTTCTT 21028

     160 lnCysAlaGlyGlyAlaAlaMetSer.....................Lys 169
         ::          ::::::::::::|||                    :::
  21027 CAGAAAAGGATGCTGTTGGCCTTTCTTACCTTGGCCTAATACCAACTGAA 20978

     170 HisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLy 186
            :::      :::   :::   :::|||      :::||||||:::::
  20977 GTTGTCTGCTGAAAGAATGTTCTCCTGTGGTTTAAGGAGCTAGAATTGAA 20928

     186 sSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCys.............. 198
         :   :::::::|||   :::   :::   :::   |||
  20927 TATCAGGTCTTTATTGTCTCCCAAAGTCCATCGGTGTACATTTACTGCTC 20878

     199 ..............GlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPhe 210
                       ::::::::|||   :::   |||:::|||::::::::::
  20877 ATTTGTTAAATAGAGAGCACTGGACTAAGACAGGGCAGAACCGGGGNATA 20828

     211 GlyLeu.............................................. 212
         |||:::
  20827 GGAATACCGAAGGGSCCGYAAGAGGGCTTGGAGACTTCACCCCATCCACC 20778

     213 .............GlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrG 225
                     :::|||      |||   |||::::::|||   ::::
```

```
  20777 CACCAAGAGCAAAAGCCTCCTTTGAACAGAGGCTCTTAACCTAAGGTCCC 20728

    225 lnGlyPheArgLysTrpGlyAlaGlnPheSerGluThrAsnSer..... 239
        ::|||           |||:::|||   |||:::   :::|||
  20727 ATGGAACAGAT......GGTGTTCAGGGATCCCAAGAGTCCTCCGAAATC 20684

    240 ............IleCysArgHis......ValLeuProLys 247
                  :::|||  |||      |||||||||:::
  20683 GTTGACAGAAACGTGTGTGCACACCTGGCGGTGCTCCCTCGG 20642

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


alignment_scores:
            Quality:   86.00            Length:    183
              Ratio:  1.075              Gaps:     11
 Percent Similarity:  43.716   Percent Identity: 24.044

alignment_block:
 US-09-674-779-2 x US-09-103-840A-2/rev   ..

 Align seg 1/1  to reverse of: US-09-103-840A-2  from: 1  to: 4403765

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaPro...IleProTh 26
         ||||||   |||   ::::::|||:::::   ||||||   :::
  4296417ACCTTGACCTCCCCGATCGTGGTCTCCGCCATAGCACCCGACGTCACCAT 4296368

      26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
          |||      ::::::|||         ||||||
  4296367CAACGAGCTGATCGATCCGGCAGTGTTGCCGTCG.............. 4296334

      43 spLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHis 59
          :::|||   |||:::::   :::             ::::::|||
  4296333..GAAATCCGCGATCGCTATTTGCGC.............ATCGACCAC 4296301

      60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
          |||                 ::::::|||||||||||||   |||
  4296300CGCGGC.................AGCTATCTGCAGATGCATTTTGCGCT 4296269

      76 rLysGlnGlu.........ValValArgTyrGlnAlaTyr.......... 86
          |||           ::::::   |||||||||
  4296268CGCGCAGCCACCGGCCTTCGCGGCGCCCTACCAGGCACTCAACGATCCGA 4296219

      87 ..LeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu 102
          :::|||:::   :::|||   :::      |||   :::
  4296218GCATGCAGGCATCGATGGGCATTTTCTGTACACCGGAGCAG......... 4296178

     103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119

          :::   ::::::|||:::   |||
  4296177...GTTCAGCAGCAGTGGGAGGATTGC..................... 4296154

     119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136

  4296154..................................................... 4296154

     136 spLeuLysSerArgGlyIleLeuProAlaAsn...........ThrGln 148
                 |||||||||:::||||||:::              |||
  4296153........CGGCGTGGGATCGTCCCGGCCGATCCGACCGTGGTGTTGCAG 4296112

     149 IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAl 165
         |||   |||::::::   :::|||:::|||      ::::::      ||
  4296111ATCCCGTCGCTGCAT...GATCCCAGCCTGGCCCCGGCGGGTAAGCAGGC 4296065

     165 aAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
         ||||                    ||||||:::   :::|||   |||
  4296064CGCG.................TCGGCGTTCGCAATGTGGTTCCCG 4296037

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


alignment_scores:
            Quality:   86.00            Length:    183
              Ratio:  1.075              Gaps:     11
 Percent Similarity:  43.716   Percent Identity: 24.044

alignment_block:
 US-09-674-779-2 x US-09-103-840A-1/rev   ..

 Align seg 1/1  to reverse of: US-09-103-840A-1  from: 1  to: 4411529

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaPro...IleProTh 26
         ||||||   |||   ::::::|||:::::   ||||||   :::
  4304162ACCTTGACCTCCCCGATCGTGGTCTCCGCCATAGCACCCGACGTCACCAT 4304113

      26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
          |||      ::::::|||         ||||||
  4304112CAACGAGCTGATCGATCCGGCAGTGTTGCCGTCG.............. 4304079

      43 spLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHis 59
          :::|||   |||:::::   :::             ::::::|||
  4304078..GAAATCCGCGATCGCTATTTGCGC.............ATCGACCAC 4304046

      60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
          |||                 ::::::|||||||||||||   |||
  4304045CGCGGC.................AGCTATCTGCAGATGCATTTTGCGCT 4304014

      76 rLysGlnGlu.........ValValArgTyrGlnAlaTyr.......... 86
          |||           ::::::   |||||||||
  4304013CGCGCAGCCACCGGCCTTCGCGGCGCCCTACCAGGCACTCAACGATCCGA 4303964
```

```
      87 ..LeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu 102
         :::|||:::   :::|||   :::      |||   :::
4303963GCATGCAGGCATCGATGGGCATTTTCTGTACACCGGAGCAG..... 4303923

     103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
           :::   ::::::|||:::   |||
4303922...GTTCAGCAGCAGTGGGAGGATTGC..... 4303899

     119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136

4303899..... 4303899

     136 spLeuLysSerArgGlyIleLeuProAlaAsn.....ThrGln 148
                 |||||||||:::||||||:::          |||
4303898........CGGCGTGGGATCGTCCCGGCCGATCCGACCGTGGTGTTGCAG 4303857

     149 IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAl 165
         |||   |||:::::::   :::|||:::|||      ::::::      ||
4303856ATCCCGTCGCTGCAT...GATCCCAGCCTGGCCCCGGCGGGTAAGCAGGC 4303810

     165 aAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
         ||||                    ||||||:::   :::|||   |||
4303809CGCG.....TCGGCGTTCGCAATGTGGTTCCCG 4303782
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-706-872-2

```
seq_documentation_block:
; Sequence 2, Application US/07706872
; Patent No. 5237056
;  GENERAL INFORMATION:
;    APPLICANT:  Fischbach, Gerald D.
;    TITLE OF INVENTION:  Identification of a Protein Which
;    TITLE OF INVENTION:  Promotes the Synthesis of Acetylcholine Receptors and Uses
;    TITLE OF INVENTION:  Therefor
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/706,872
;      FILING DATE:  19910529
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  HU91-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2188 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  172..972
US-07-706-872-2


alignment_scores:
```

```
            Quality:   85.50            Length:    207
              Ratio:  0.830              Gaps:     12
Percent Similarity:  49.758   Percent Identity: 24.638

alignment_block:
 US-09-674-779-2 x US-07-706-872-2   ..

 Align seg 1/1  to: US-07-706-872-2  from: 1  to: 2188

      21 SerAlaProIleProThrAsnProGlnValSerProIleLysThrProSe 37
         |||:::|||:::|||   :::      |||:::                ::
     691 AGCAGCCCCGTGCCACAGGACGTCTTCGTGGCCGATTGCTTT......AA 734

      37 rValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisA 54
         ::::   :::|||:::   :::||||||                ::   :
     735 CATCACAGTGACTGAGTACAGCATTGGCCCTGCTGCCAAGAAGAACACCT 784

      54 spGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeu 70
         ::|||:::|||:::              |||      :::|||   :::
     785 CCGAGGCTGTGGCG.............GCAGCAAACCAAACGGAGGTG 819

      71 GlnMetHisHisAlaThrLysGlnGluValValArg.....TyrGl 84
         :::|||:::::   :::      :::|||:::|||          ||
     820 GAGATGGAGAACAAAGTGGTGACGAAGGTGATCCGCGAGATGTGCGTGCA 869

      84 nAlaTyrLeuGlnSerArgLeuGlyAsn..... 93
         |   |||   :::   ||||||:::::
     870 GCAGTACCGCGAGTACCGCCTGGCCTCGGGCATCCAGCTGCACCCTGCTG 919

      94 .....TyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSer... 107
           :::|||      :::      ||||||||||||    :::
     920 ACACCTGGCTCGCCGTCCTCCTCCTCCTCCTCACCACCCTTTTTGCCATG 969

     108 ......TrpGlnAla.....CysGlyHisGluProTyrGlnLe 118
           |||:::|||          ||||||   |||      :::
     970 CACTGATGGGATGCCGTGCCCCGGCCCTGTGGCAGTGAGATGACATCGTG 1019

     118 uProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrG 135
          |||   |||   ||||||      ||||||      :::::
    1020 TCCCCGTGCCCACCCATGGGGT......GTTCCTTGTCCTCGCTTTTGTC 1063

     135 lnAspLeu......LysSerArgGlyIleLeuProAlaAsnThrGlnIle 149
           :::|||         |||         |||||||||:::::::::::
    1064 CATCTTTGGTGAAGATGTCCCCCCGCTGCCTCCCCGCAGGCTCTGATTTG 1113

     150 .........ArgSerValTyrArgAsnProGluLeuAsnGlnCysAla.. 162
                 :::::::::         |||      ::::::   |||
    1114 GGCAAATGGGAGGGGATTTTGTCCTGTCCTGGTCGTGGCAGGACGGCTGC 1163

     163 ....GlyGlyAlaAlaMetSerLysHis.....................L 171
             ||||||::::::      |||:::                     |
    1164 TGGTGGTGGAGTGGGATGCCCAAAAAATGGCCTTCACCACTTCCTCCTCC 1213

     171 euThrAsnSerAlaIleAspIleTrpVal...ProAspLeuGluIleLys 186
         ||      |||:::   :::||||||:::   |||   |||   :::
    1214 TCTTCCTTTCTGGGGCGGAGATATGGGCTCGTCCAGCCCTTATTGTCCCT 1263

     187 SerGlnAlaLeuTyrGluLeu 193
         ::::::|||      ::::::
    1264 GCAAGAGCGTATCTGAAAATC 1284

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-673-312-9

seq_documentation_block:
; Sequence 9, Application US/08673312
; Patent No. 5891699
;  GENERAL INFORMATION:
;    APPLICANT:  BOULAIN, JEAN-CLAUDE
;    APPLICANT:  CATTOLICO, LAURENCE
;    APPLICANT:  DUCANCEL, FREDERIC
;    APPLICANT:  MENEZ, ANDRE
;    TITLE OF INVENTION:  MODIFIED BACTERIAL ALKALINE PHOSPHATASES
;    TITLE OF INVENTION:  AND THEIR APPLICATION
```

```
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/673,312
;      FILING DATE:  28-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 95-07833
;      FILING DATE:  29-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  846-380-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6162 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-673-312-9


alignment_scores:
            Quality:   85.00            Length:     59
              Ratio:  2.179              Gaps:      0
 Percent Similarity:  66.102   Percent Identity: 30.508

alignment_block:
 US-09-674-779-2 x US-08-673-312-9/rev   ..

 Align seg 1/1  to reverse of: US-08-673-312-9  from: 1  to: 6162

      32 ProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHi 48
         |||   :::|||   :::||||||:::      ||||||   |||:::||
    2766 CCATATTCAACGGGAAACGTCTTGCTCGAGGGTGACAAGGCAGGAAACCA 2717

      48 sHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaH 65
         ||||   |||:::      ::::::   :::   ::::::          :
    2716 CCATCAACACCAGACCCGCGACAAACAGACGGGGGCCGCCTGCCCCCATC 2667

      65 isPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValVal 81
         :::::::|||      ::::::|||::::::   :::|||:::|||
    2666 AAATTCAGACACATAATCATCAACATCGCGGTGGAAGGCAGCACGTTTAC 2617

      82 ArgTyrGlnAlaTyrLeuGlnSerArg 90
            :::   ||||||   :::||||||
    2616 GCCCATTACGCCTACCCACATTCCCGC 2590

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-276-099A-1

seq_documentation_block:
; Sequence 1, Application US/08276099A
; Patent No. 5591825
;  GENERAL INFORMATION:
;    APPLICANT:  McKnight, Steven L
;    APPLICANT:  Hou, Jinzhao
;    TITLE OF INVENTION:  INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
;    TITLE OF INVENTION:  BINDING ASSAYS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,099A
;      FILING DATE:  15-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59451-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3046 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  166..2706
US-08-276-099A-1


alignment_scores:
            Quality:   83.50            Length:    118
              Ratio:  1.392              Gaps:      7
 Percent Similarity:  50.847   Percent Identity: 29.661

alignment_block:
 US-09-674-779-2 x US-08-276-099A-1/rev   ..

 Align seg 1/1  to reverse of: US-08-276-099A-1  from: 1  to: 3046

      66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValAr 82
         |||:::||||||              :::   ||||||   |||:::||
    1016 TTCAGTACCTGGGGGGGCTGCTTCTCCACCAGGAAGCAACTGGTGACGAG 967

      82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetS 99
         |                 :::||||||                 |||:::|
     966 GGTTCTCAGGACTTCATCCAGCCGG.................CCAGTCA 935

      99 erGlnLeuLeuThrThrAlaArg.SerTrpGlnAlaCysGlyHisGluPr 115
         ||:::            ||||||  ||||||      :::   |||:::||
     934 GCGAT............GCCCGGGTCTTGGGCTCAAGCTCCCCACCAGCC 897

     115 oTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuH 132
         |      |||||||||           ::::::      |||   :::|
     896 G...CCCCTACCTCCTGCTGTAGCTGGGAATAAATGTCCACCAGGCTTTC 850

     132 isLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGln 148
         ||                |||   |||:::   |||::::::   |||
     849 AC..............ACCTCTCCTGGAGTGGGGCCAGGCTCTCCTCAA 815

     149 ...IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGl 164
            :::|||                    ::::::      :::::
     814 ACGGTGCGC.....................CATTCCCTGCCAGCTGC 789
```

```
     164 yAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValP 181
         :||||||:::||||||      ::::::||||||:::   :::|||:::|
     788 TGCTGCCGTTTCCAAATCTGGATCCTCTTCAGCACTAGGGCTTTGGCTGC 739

     181 ro 181
         ||
     738 CT 737

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-781-890-1

seq_documentation_block:
; Sequence 1, Application US/08781890
; Patent No. 5710266
;  GENERAL INFORMATION:
;    APPLICANT:  McKnight, Steven L
;    APPLICANT:  Hou, Jinzhao
;    TITLE OF INVENTION:  INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
;    TITLE OF INVENTION:  BINDING ASSAYS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,890
;      FILING DATE:  05-JAN-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/276,099
;      FILING DATE:  15-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59451-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3046 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  166..2706
US-08-781-890-1


alignment_scores:
            Quality:   83.50            Length:    118
              Ratio:  1.392              Gaps:      7
 Percent Similarity:  50.847   Percent Identity: 29.661

alignment_block:
 US-09-674-779-2 x US-08-781-890-1/rev   ..

 Align seg 1/1  to reverse of: US-08-781-890-1  from: 1  to: 3046

      66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValAr 82
         |||:::||||||              :::   ||||||   |||:::||
    1016 TTCAGTACCTGGGGGGGCTGCTTCTCCACCAGGAAGCAACTGGTGACGAG 967

      82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetS 99
         |              :::|||||                  |||:::|
     966 GGTTCTCAGGACTTCATCCAGCCGG.................CCAGTCA 935

      99 erGlnLeuLeuThrThrAlaArg.SerTrpGlnAlaCysGlyHisGluPr 115
         ||:::           ||||||  |||||      :::   |||:::||
     934 GCGAT...........GCCCGGGTCTTGGGCTCAAGCTCCCCACCAGCC 897

     115 oTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuH 132
         |     |||||||||           ::::::     |||   :::|
     896 G...CCCCTACCTCCTGCTGTAGCTGGGAATAAATGTCCACCAGGCTTTC 850

     132 isLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGln 148
         ||                  |||   |||:::   |||::::::   |||
     849 AC..............ACCTCTCCTGGAGTGGGGCCAGGCTCTCCTCAA 815

     149 ...IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGl 164
            :::|||                          ::::::      :::::
     814 ACGGTGCGC.....................CATTCCCTGCCAGCTGC 789

     164 yAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValP 181
         :||||||:::||||||      ::::::||||||:::   :::|||:::|
     788 TGCTGCCGTTTCCAAATCTGGATCCTCTTCAGCACTAGGGCTTTGGCTGC 739

     181 ro 181
         ||
     738 CT 737

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-087-465-11

seq_documentation_block:
; Sequence 11, Application US/09087465A
; Patent No. 6160092
; GENERAL INFORMATION:
;  APPLICANT: Vinkemeier, Uwe
;  APPLICANT: Chen, Xiaomin
;  APPLICANT: Darnell Jr., James E
;  APPLICANT: Kuriyan, John
;  TITLE OF INVENTION: A CRYSTAL OF THE CORE PORTION OF A STAT AND METHODS OF
;  TITLE OF INVENTION: USE
;  FILE REFERENCE: 600-1-229
;  CURRENT APPLICATION NUMBER: US/09/087,465A
;  CURRENT FILING DATE: 1998-05-29
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 3046
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-087-465-11


alignment_scores:
            Quality:   83.50            Length:    118
              Ratio:  1.392              Gaps:      7
 Percent Similarity:  50.847   Percent Identity: 29.661

alignment_block:
 US-09-674-779-2 x US-09-087-465-11/rev   ..

 Align seg 1/1  to reverse of: US-09-087-465-11  from: 1  to: 3046

      66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValAr 82
         |||:::||||||              :::   ||||||   |||:::||
    1016 TTCAGTACCTGGGGGGGCTGCTTCTCCACCAGGAAGCAACTGGTGACGAG 967

      82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetS 99
         |              :::|||||                  |||:::|
     966 GGTTCTCAGGACTTCATCCAGCCGG.................CCAGTCA 935

      99 erGlnLeuLeuThrThrAlaArg.SerTrpGlnAlaCysGlyHisGluPr 115
         ||:::           ||||||  |||||      :::   |||:::||
     934 GCGAT...........GCCCGGGTCTTGGGCTCAAGCTCCCCACCAGCC 897
```

```
    115 oTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuH 132
        |      |||||||||            ::::::      |||   :::|
    896 G...CCCCTACCTCCTGCTGTAGCTGGGAATAAATGTCCACCAGGCTTTC 850

    132 isLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGln 148
        ||              |||   |||:::   |||::::::   |||
    849 AC..............ACCTCTCCTGGAGTGGGGCCAGGCTCTCCTCAA 815

    149 ...IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGl 164
           :::|||                    ::::::      :::::
    814 ACGGTGCGC.....................CATTCCCTGCCAGCTGC 789

    164 yAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValP 181
        :||||||:::|||||||      ::::::||||||:::   :::|||:::|
    788 TGCTGCCGTTTCCAAATCTGGATCCTCTTCAGCACTAGGGCTTTGGCTGC 739

    181 ro 181
        ||
    738 CT 737

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-621-193A-1

seq_documentation_block:
; Sequence 1, Application US/07621193A
; Patent No. 5187087
;  GENERAL INFORMATION:
;    APPLICANT:  Sondermeyer, Paulus Jacobus Antonius
;    APPLICANT:  Claessens, Johannes Antonius Joseph
;    APPLICANT:  Mockett, Albert Philip Adrian
;    TITLE OF INVENTION:  Recombinant herpesvirus of
;    TITLE OF INVENTION:  turkeys and live vector vaccines derived thereof
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Organon Teknika Corporation
;      ADDRESSEE:  Biotechnology Research Institute
;      STREET:  1330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  U.S.A.
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/621,193A
;      FILING DATE:  19901130
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donna Bobrowicz
;      REGISTRATION NUMBER:  32,196
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301)258-5200
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4527 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpesvirus of turkey
;      STRAIN:  PB-THV1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..81
;      OTHER INFORMATION:  /label= end_of_ORF1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..945
;      OTHER INFORMATION:  /label= ORF2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (1084..2124)
;      OTHER INFORMATION:  /label= ORF3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (2322..3170)
;      OTHER INFORMATION:  /label= ORF4
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3320..4504
;      OTHER INFORMATION:  /label= ORF5
US-07-621-193A-1


alignment_scores:
            Quality:   82.50            Length:    201
              Ratio:  0.743              Gaps:     13
 Percent Similarity:  55.224   Percent Identity: 24.378

alignment_block:
 US-09-674-779-2 x US-07-621-193A-1   ..

 Align seg 1/1  to: US-07-621-193A-1  from: 1  to: 4527

     14 SerSerMetLeuValAlaCysSerAlaProIleProThrAsnProGlnVa 30
        |||:::   |||:::      ::::::   :::|||:::|||||||:::
   2866 TCAAACAGTCTCACAATAAATGCATCTCTCGTTCCTGCCAATCCGGAACC 2915

     30 lSer...ProIleLysThrProSerValLeuIleThr..... 41
         :::   |||:::         ::::::::::|||
   2916 GGGCATACCACTCCCACCTGCCGATTTAATTCTCACAATTGGGCGATGCC 2965

     42 ..LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57
           :::   |||   :::         :::::::::|||   ::::::|||
   2966 GGCGGGGCAAAACGAATGTGGATTTGGCAAACCGACACAGGTCTGCTGTA 3015

     58 ...SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHi 73
          :::::::::|||      :::   |||     :::
   3016 CGGACTAATATGGGCACACCCACATCATTC......TTCAGATGCTCCAT 3059

     73 sHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerA 90
         |||:::::   :::::::::   |||   :::|||      ::::::|
   3060 GCATTGTTCTATGAGAAAGATCCATAGGGTGGAGGCAGCGTCACGAGATC 3109

     90 rgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArg 106
        ||   ||||||                  :::      ::::::::::::
   3110 GCCCAGGCAAT.................CGATCGCATTCGTCTAGTAAA 3141

    107 SerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGlu..... 121
             ::::::      |||   |||         |||   |||
   3142 GTGACGAGAGTTATCATGCACACACCCATGCCCACGCCTTCCGAATAACT 3191

    122 .HisLeuTrp......GlyGlnIleValProThrLeuHisLeuTyrGlnA 136
         :::|||||      |||::::::      |||      |||
   3192 GGAGCTGTGGAAGATCGGAAACGTCTTTTTGACTGCCGGTCTCGTA.... 3237

    136 spLeuLysSerArg.GlyIleLeuProAlaAsnThrGlnIleArgSerVa 152
          |||   |||::: |||:::   |||           |||      ::
   3238 ..CTACTTTCGCACAGGTGTATACCCGGACGCGTACTATATATTTTATAT 3285

    152 lTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerL 169
        :   ::::::||||||::::::              ||||||||||:::
   3286 CATTCAACGTCCCGAAATTACATACGTG.........GCGGCGATGGAAG 3326

    169 ysHisLeuThrAsnSerAlaIle......Asp.IleTrpValProAspLe 183
           :::   :::|||   :::       ::: |||||||||   :::::
   3327 TAGATGTTGAGTCTTCGAAAGTAAGTGCCTCGAATATGGGTATTGTCTGT 3376

    183 uGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGln 199
```

```
        ::::|||   |||:::|||:::   :::|||||||||||||   |||:::
    3377 GAAAATA...TCGAAAGCGGTACGACGGTTGCAGAACCGTCGATGTCGC 3422

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-018-489C-1

seq_documentation_block:
; Sequence 1, Application US/08018489C
; Patent No. 5470734
;  GENERAL INFORMATION:
;    APPLICANT:  Sondermeyer, Paulus Jacobus Antonius
;    APPLICANT:  Claessens, Johannes Antonius Joseph
;    APPLICANT:  Mockett, Albert Philip Adrian
;    TITLE OF INVENTION:  Recombinant herpesvirus of
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Akzo Pharma
;      STREET:  1330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  U.S.A.
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/018,489C
;      FILING DATE:  February 12, 1993
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/621,193
;      FILING DATE:  30-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donna Bobrowicz
;      REGISTRATION NUMBER:  32,196
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301)258-5200
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4527 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpesvirus of turkey
;      STRAIN:  PB-THV1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..81
;      OTHER INFORMATION:  /label= end_of_ORF1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..945
;      OTHER INFORMATION:  /label= ORF2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (1084..2124)
;      OTHER INFORMATION:  /label= ORF3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (2322..3170)
;      OTHER INFORMATION:  /label= ORF4
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3320..4504
;      OTHER INFORMATION:  /label= ORF5
US-08-018-489C-1


alignment_scores:
           Quality:   82.50            Length:    201
              Ratio:  0.743                Gaps:     13
 Percent Similarity:  55.224   Percent Identity: 24.378

alignment_block:
 US-09-674-779-2 x US-08-018-489C-1   ..

 Align seg 1/1  to: US-08-018-489C-1  from: 1  to: 4527

      14 SerSerMetLeuValAlaCysSerAlaProIleProThrAsnProGlnVa 30
         |||:::   |||:::      ::::::   :::|||:::|||||||:::
    2866 TCAAACAGTCTCACAATAAATGCATCTCTCGTTCCTGCCAATCCGGAACC 2915

      30 lSer...ProIleLysThrProSerValLeuIleThr..............  41
          :::   |||:::          ::::::::::::|||
    2916 GGGCATACCACTCCCACCTGCCGATTTAATTCTCACAATTGGGCGATGCC 2965

      42 ..LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57
           :::   |||   :::         :::::::::|||   :::::|||
    2966 GGCGGGGCAAAACGAATGTGGATTTGGCAAACCGACACAGGTCTGCTGTA 3015

      58 ...SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHi 73
            :::::::::|||      :::   |||      :::
    3016 CGGACTAATATGGGCACACCCACATCATTC......TTCAGATGCTCCAT 3059

      73 sHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerA 90
          |||::::::   :::::::::   |||   :::|||      ::::::|
    3060 GCATTGTTCTATGAGAAAGATCCATAGGGTGGAGGCAGCGTCACGAGATC 3109

      90 rgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArg 106
          ||   ||||||                  :::      ::::::::::::
    3110 GCCCAGGCAAT.................CGATCGCATTCGTCTAGTAAA 3141

     107 SerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGlu..... 121
           ::::::      |||   |||         |||   |||
    3142 GTGACGAGAGTTATCATGCACACACCCATGCCCACGCCTTCCGAATAACT 3191

     122 .HisLeuTrp......GlyGlnIleValProThrLeuHisLeuTyrGlnA 136
          :::||||||      |||::::::      |||      |||
    3192 GGAGCTGTGGAAGATCGGAAACGTCTTTTTGACTGCCGGTCTCGTA.... 3237

     136 spLeuLysSerArg.GlyIleLeuProAlaAsnThrGlnIleArgSerVa 152
           |||   |||::: |||:::   |||         |||      ::
    3238 ..CTACTTTCGCACAGGTGTATACCCGGACGCGTACTATATATTTTATAT 3285

     152 lTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerL 169
          :   ::::::||||||::::::               ||||||||||:::
    3286 CATTCAACGTCCCGAAATTACATACGTG.........GCGGCGATGGAAG 3326

     169 ysHisLeuThrAsnSerAlaIle......Asp.IleTrpValProAspLe 183
           :::   :::|||   :::      ::: ||||||||   :::::
    3327 TAGATGTTGAGTCTTCGAAAGTAAGTGCCTCGAATATGGGTATTGTCTGT 3376

     183 uGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGln 199
         ::::|||   |||:::|||:::   :::|||||||||||   |||:::
    3377 GAAAATA...TCGAAAGCGGTACGACGGTTGCAGAACCGTCGATGTCGC 3422

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-23

seq_documentation_block:
; Sequence 23, Application US/08339214
; Patent No. 6348334
;  GENERAL INFORMATION:
;    APPLICANT:  Nagata, Shigikazu
;    APPLICANT:  Suda, Takashi
;    APPLICANT:  Takahashi, Tomoniro
;    APPLICANT:  Nakamura, No. 6348334io
;    TITLE OF INVENTION:  A Fas Ligand, A Fragment Thereof and DNA
;    TITLE OF INVENTION:  Encoding the Same
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
```

```
;       STATE:  Virginia
;       COUNTRY:  USA
;       ZIP:  22040-0747
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/339,214
;       FILING DATE:  10-NOV-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Murphy Jr., Gerald M.
;       REGISTRATION NUMBER:  28,977
;       REFERENCE/DOCKET NUMBER:  1110-139P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-205-8000
;       TELEFAX:  703-205-8050
;   INFORMATION FOR SEQ ID NO:  23:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  837 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  cDNA to mRNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  mouse
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..837
US-08-339-214-23


alignment_scores:
            Quality:   82.00              Length:    253
              Ratio:  0.837                Gaps:     16
 Percent Similarity:  38.735   Percent Identity: 25.296

alignment_block:
 US-09-674-779-2 x US-08-339-214-23   ..

 Align seq 1/1  to: US-08-339-214-23  from: 1  to: 837

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSer..... 37
         |||   |||       |||   |||||||||::::      ||||||
     133 CCGCCACCTCCACCACCACCTGTGTCACCACTACCACCGCCATCACAACC 182

      38 ..........ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
                       :::|||      |||      ||||||:::|||:
     183 ACTCCCACTGCCGCCACTGACCCCTCTAAAGAAGAAGGACCACAACACAA 232

      51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
         ::
     233 AT..................................................... 234

      68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
            ||||||                          ||||||   :::::
     235 CTGTGGCTA.....................CCGGTGGTATTTTTCAT 260

      84 n.........AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         :         |||   :::         ||||||   |||
     261 GGTTCTGGTGGCTCTGGTTGGAATGGGATTAGGAATGTAT.......... 300

      98 etSerGlnLeuLeu.ThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
               ||||||::: |||      |||:::||||||           ::
     301 .....CAGCTCTTCCACCTGCAGAAGGAACTGGCAGAACTCCGTGAGTTC 345

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
         :|||   :::           ||||||      :::      ||||||
     346 ACCAACCAAAGCCTTAAAGTATCATCTTTTGAAAAGCAAATAGCCAACCC 395

     131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
           :::|||      :::      ||||||||||||:::   |||    ||
     396 CAGTACACCCTCTGAAAAAA...AAGAGCCGAGGAGTGTGGCCCATTTAA 442

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
         |                  |||:::|||   |||:::          :
     443 C.....................AGGGAACCCCCACTCAAGGTCCATCCCT 471

     163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
         ::|||      |||::::::|||:::   |||   :::   |||:::|||
     472 CTGGA......ATGGGAAGACACATATGGAACCGCTCT...GATCTCTGG 512

     179 ..................................................... 179

     513 AGTGAAGTATAAGAAAGGTGGCCTTGTGATCAACGAAACTGGGTTGTACT 562

     180 ..................ValProAspLeuGluIle............. 185
                            :::|||         :::
     563 TCGTGTATTCCAAAGTATACTTCCGGGGTCAGTCTTGCAACAACCAGCCC 612

     186 ..LysSerGlnAlaLeuTyrGlu..........................  192
           |||   |||:::||||||||
     613 CTAAACCACAAGGTCTATATGAGGAACTCTAAGTATCCTGAGGATCTGGT 662

     193 ............LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGl 205
                       :::   :::|||   :::|||||||   :::   ||
     663 GCTAATGGAGGAGAAGAGGTTGAACTACTGCACTACTGGACAGATATGGG 712

     205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
         |   :::|||    :::|||      :::      |||:::|||   |||
     713 CCCACAGCAGCTACCTGGGGGCAGTATTCAATCTTACCAGTGCTGACCAT 762

     222 Leu 222
         |||
     763 TTA 765

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-36

seq_documentation_block:
; Sequence 36, Application US/08339214
; Patent No. 6348334
;  GENERAL INFORMATION:
;    APPLICANT:  Nagata, Shigikazu
;    APPLICANT:  Suda, Takashi
;    APPLICANT:  Takahashi, Tomoniro
;    APPLICANT:  Nakamura, No. 6348334io
;    TITLE OF INVENTION:  A Fas Ligand, A Fragment Thereof and DNA
;    TITLE OF INVENTION:  Encoding the Same
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/339,214
;      FILING DATE:  10-NOV-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1110-139P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
```

```
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  YES
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..32
;    FEATURE:
;      NAME/KEY:  3'UTR
;      LOCATION:  872..927
US-08-339-214-36


alignment_scores:
            Quality:   82.00              Length:    253
              Ratio:  0.837                Gaps:     16
 Percent Similarity:  38.735   Percent Identity: 25.296

alignment_block:
 US-09-674-779-2 x US-08-339-214-36   ..

 Align seg 1/1  to: US-08-339-214-36  from: 1  to: 927

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSer..... 37
         |||   |||       |||   |||||||||::       ||||||
     165 CCGCCACCTCCACCACCACCTGTGTCACCACTACCACCGCCATCACAACC 214

      38 ..........ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
                      :::|||      |||       ||||||:::|||:
     215 ACTCCCACTGCCGCCACTGACCCCTCTAAAGAAGAAGGACCACAACACAA 264

      51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
         ::
     265 AT..................................................... 266

      68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
            ||||||                          ||||||   :::::
     267 CTGTGGCTA.......................CCGGTGGTATTTTTCAT 292

      84 n.........AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         :        |||   :::          ||||||   |||
     293 GGTTCTGGTGGCTCTGGTTGGAATGGGATTAGGAATGTAT.......... 332

      98 etSerGlnLeuLeu.ThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
            |||||||:: |||      |||:::||||||          ::
     333 .....CAGCTCTTCCACCTGCAGAAGGAACTGGCAGAACTCCGTGAGTTC 377

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
         :|||   :::             ||||||      :::      ||||||
     378 ACCAACCAAAGCCTTAAAGTATCATCTTTTGAAAAGCAAATAGCCAACCC 427

     131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
           :::|||      :::      ||||||||||||::   |||    ||
     428 CAGTACACCCTCTGAAAAAA...AAGAGCCGAGGAGTGTGGCCCATTTAA 474

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
         |                    |||:::|||   |||:::         :
     475 C...................AGGGAACCCCCACTCAAGGTCCATCCCT 503

     163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
         ::|||      |||::::::|||:::   |||   :::   |||:::|||
     504 CTGGA......ATGGGAAGACACATATGGAACCGCTCT...GATCTCTGG 544

     179 ..................................................... 179

     545 AGTGAAGTATAAGAAAGGTGGCCTTGTGATCAACGAAACTGGGTTGTACT 594

     180 ..................ValProAspLeuGluIle............. 185

                                  :::|||         :::
     595 TCGTGTATTCCAAAGTATACTTCCGGGGTCAGTCTTGCAACAACCAGCCC 644

     186 ..LysSerGlnAlaLeuTyrGlu..........................  192
           |||   |||:::||||||||||
     645 CTAAACCACAAGGTCTATATGAGGAACTCTAAGTATCCTGAGGATCTGGT 694

     193 ............LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGl 205
                        :::   :::|||   :::|||||||   :::   ||
     695 GCTAATGGAGGAGAAGAGGTTGAACTACTGCACTACTGGACAGATATGGG 744

     205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
         |   :::|||    :::|||      :::      |||:::|||   |||
     745 CCCACAGCAGCTACCTGGGGGCAGTATTCAATCTTACCAGTGCTGACCAT 794

     222 Leu 222
         |||
     795 TTA 797

 seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-103

 seq_documentation_block:
 ; Sequence 103, Application US/08339214
 ; Patent No. 6348334
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Nagata, Shigikazu
 ;    APPLICANT:  Suda, Takashi
 ;    APPLICANT:  Takahashi, Tomoniro
 ;    APPLICANT:  Nakamura, No. 6348334io
 ;    TITLE OF INVENTION:  A Fas Ligand, A Fragment Thereof and DNA
 ;    TITLE OF INVENTION:  Encoding the Same
 ;    NUMBER OF SEQUENCES:  103
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
 ;      STREET:  P.O. Box 747
 ;      CITY:  Falls Church
 ;      STATE:  Virginia
 ;      COUNTRY:  USA
 ;      ZIP:  22040-0747
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/08/339,214
 ;      FILING DATE:  10-NOV-1994
 ;      CLASSIFICATION:  435
 ;    ATTORNEY/AGENT INFORMATION:
 ;      NAME:  Murphy Jr., Gerald M.
 ;      REGISTRATION NUMBER:  28,977
 ;      REFERENCE/DOCKET NUMBER:  1110-139P
 ;    TELECOMMUNICATION INFORMATION:
 ;      TELEPHONE:  703-205-8000
 ;      TELEFAX:  703-205-8050
 ;  INFORMATION FOR SEQ ID NO:  103:
 ;    SEQUENCE CHARACTERISTICS:
 ;      LENGTH:  927 base pairs
 ;      TYPE:  nucleic acid
 ;      STRANDEDNESS:  not relevant
 ;      TOPOLOGY:  not relevant
 ;    MOLECULE TYPE:  cDNA to rRNA
 ;    HYPOTHETICAL:  NO
 ;    ORIGINAL SOURCE:
 ;      ORGANISM:  mouse
 ;      DEVELOPMENTAL STAGE:  rearranged
 ;    FEATURE:
 ;      NAME/KEY:  mutation
 ;      LOCATION:  replace(849, "")
 ;      OTHER INFORMATION:  /note= "T residue 849 in SEQ. ID.
 ;      OTHER INFORMATION:  36 is C in the present sequence"
 US-08-339-214-103
```

```
alignment_scores:
            Quality:   82.00              Length:    253
              Ratio:  0.837                Gaps:     16
 Percent Similarity:  38.735   Percent Identity: 25.296

alignment_block:
 US-09-674-779-2 x US-08-339-214-103   ..

 Align seg 1/1  to: US-08-339-214-103  from: 1  to: 927

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSer..... 37
         |||   |||      |||   |||||||||:::      ||||||
     165 CCGCCACCTCCACCACCACCTGTGTCACCACTACCACCGCCATCACAACC 214

      38 .........ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
                     :::|||      |||      ||||||:::|||:
     215 ACTCCCACTGCCGCCACTGACCCCTCTAAAGAAGAAGGACCACAACACAA 264

      51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
         ::
     265 AT..................................................... 266

      68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
            ||||||                          ||||||   :::::
     267 CTGTGGCTA.....................CCGGTGGTATTTTTCAT 292

      84 n.........AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         :         |||   :::         ||||||   |||
     293 GGTTCTGGTGGCTCTGGTTGGAATGGGATTAGGAATGTAT.......... 332

      98 etSerGlnLeuLeu.ThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
            |||||||:: |||      |||:::||||||           ::
     333 .....CAGCTCTTCCACCTGCAGAAGGAACTGGCAGAACTCCGTGAGTTC 377

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
         :|||   :::          ||||||      :::      ||||||
     378 ACCAACCAAAGCCTTAAAGTATCATCTTTTGAAAAGCAAATAGCCAACCC 427

     131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
           :::|||      :::      ||||||||||||:::   |||    ||
     428 CAGTACACCCTCTGAAAAAA...AAGAGCCGAGGAGTGTGGCCCATTTAA 474

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
         |                    |||:::|||   |||:::         :
     475 C.....................AGGGAACCCCCACTCAAGGTCCATCCCT 503

     163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
         ::|||      |||::::::|||:::   |||   :::   |||:::|||
     504 CTGGA......ATGGGAAGACACATATGGAACCGCTCT...GATCTCTGG 544

     179 ..................................................... 179

     545 AGTGAAGTATAAGAAAGGTGGCCTTGTGATCAACGAAACTGGGTTGTACT 594

     180 ...................ValProAspLeuGluIle.............. 185
                            :::|||         :::
     595 TCGTGTATTCCAAAGTATACTTCCGGGGTCAGTCTTGCAACAACCAGCCC 644

     186 ..LysSerGlnAlaLeuTyrGlu.......................... 192
           |||   |||:::|||||||||
     645 CTAAACCACAAGGTCTATATGAGGAACTCTAAGTATCCTGAGGATCTGGT 694

     193 ............LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGl 205
                      :::   :::|||   :::||||||   :::   ||
     695 GCTAATGGAGGAGAAGAGGTTGAACTACTGCACTACTGGACAGATATGGG 744

     205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
         |   :::|||    :::|||      :::     |||:::|||   |||
     745 CCCACAGCAGCTACCTGGGGGCAGTATTCAATCTTACCAGTGCTGACCAT 794

     222 Leu 222
         |||
```

```
     795 TTA 797

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-31

seq_documentation_block:
; Sequence 31, Application US/08339214
; Patent No. 6348334
;  GENERAL INFORMATION:
;    APPLICANT:  Nagata, Shigikazu
;    APPLICANT:  Suda, Takashi
;    APPLICANT:  Takahashi, Tomoniro
;    APPLICANT:  Nakamura, No. 6348334io
;    TITLE OF INVENTION:  A Fas Ligand, A Fragment Thereof and DNA
;    TITLE OF INVENTION:  Encoding the Same
;    NUMBER OF SEQUENCES:  103
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/339,214
;      FILING DATE:  10-NOV-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1110-139P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1707 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;    IMMEDIATE SOURCE:
;      CLONE:  LambdaMFL5, LambdaMFL18
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  125..961
;      OTHER INFORMATION:  /label= Figs_23-24
US-08-339-214-31


alignment_scores:
            Quality:   82.00              Length:    253
              Ratio:  0.837                Gaps:     16
 Percent Similarity:  38.735   Percent Identity: 25.296

alignment_block:
 US-09-674-779-2 x US-08-339-214-31   ..

 Align seg 1/1  to: US-08-339-214-31  from: 1  to: 1707

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSer..... 37
         |||   |||      |||   |||||||||:::      ||||||
     257 CCGCCACCTCCACCACCACCTGTGTCACCACTACCACCGCCATCACAACC 306

      38 .........ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
```

```
                    :::|||      |||      ||||||:::|||:
    307 ACTCCCACTGCCGCCACTGACCCCTCTAAAGAAGAAGGACCACAACACAA 356

     51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
        ::
    357 AT..................................................... 358

     68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
         ||||||                          ||||||   :::::
    359 CTGTGGCTA.....................CCGGTGGTATTTTTCAT 384

     84 n.........AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
        :         |||   :::         ||||||   |||
    385 GGTTCTGGTGGCTCTGGTTGGAATGGGATTAGGAATGTAT.......... 424

     98 etSerGlnLeuLeu.ThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
           ||||||:: |||      |||:::||||||              ::
    425 .....CAGCTCTTCCACCTGCAGAAGGAACTGGCAGAACTCCGTGAGTTC 469

    114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
        :|||   :::             ||||||      :::      ||||||
    470 ACCAACCAAAGCCTTAAAGTATCATCTTTTGAAAAGCAAATAGCCAACCC 519

    131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
          :::|||      :::      ||||||||||||::   |||    ||
    520 CAGTACACCCTCTGAAAAAA...AAGAGCCGAGGAGTGTGGCCCATTTAA 566

    146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
        |                     |||:::|||   |||:::         :
    567 C...................AGGGAACCCCCACTCAAGGTCCATCCCT 595

    163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
        ::|||      |||:::::|||:::   |||   :::   |||:::|||
    596 CTGGA......ATGGGAAGACACATATGGAACCGCTCT...GATCTCTGG 636

    179 ..................................................... 179

    637 AGTGAAGTATAAGAAAGGTGGCCTTGTGATCAACGAAACTGGGTTGTACT 686

    180 ..................ValProAspLeuGluIle............. 185
                          :::|||      :::
    687 TCGTGTATTCCAAAGTATACTTCCGGGGTCAGTCTTGCAACAACCAGCCC 736

    186 ..LysSerGlnAlaLeuTyrGlu........................ 192
          |||   |||:::|||||||||
    737 CTAAACCACAAGGTCTATATGAGGAACTCTAAGTATCCTGAGGATCTGGT 786

    193 ............LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGl 205
                    :::   :::|||   :::||||||   :::   ||
    787 GCTAATGGAGGAGAAGAGGTTGAACTACTGCACTACTGGACAGATATGGG 836

    205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
        |   :::|||    :::|||      :::       |||:::|||   |||
    837 CCCACAGCAGCTACCTGGGGGCAGTATTCAATCTTACCAGTGCTGACCAT 886

    222 Leu 222
        |||
    887 TTA 889

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-038-832-3

seq_documentation_block:
; Sequence 3, Application US/09038832
; Patent No. 6146845
;  GENERAL INFORMATION:
;    APPLICANT:  KIKLY, KRISTINE
;    APPLICANT:  ERICKSON-MILLER, CONNIE
;    TITLE OF INVENTION:  Sialoadhesin Family Member-2
;    TITLE OF INVENTION:  (SAF-2)
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/038,832
;      FILING DATE:  11-MAR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/041,886
;      FILING DATE:  02-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-50018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1597 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-038-832-3


alignment_scores:
            Quality:   81.50              Length:    229
              Ratio:  0.849                Gaps:      9
 Percent Similarity:  41.921   Percent Identity: 22.271

alignment_block:
 US-09-674-779-2 x US-09-038-832-3   ..

 Align seg 1/1  to: US-09-038-832-3  from: 1  to: 1597

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
         |||   |||   :::|||:::|||   |||::::::   |||::::::||
     677 CCTCATCCTAGGGACCCTAGAGTCTGGCCACTCCAGGAACCTGACCTGCT 726

      39 uIleThrLysAspLys...IleGlyAspHisHisThrHisGluHisAspG 55
         |   :::                |||||               ||||||
     727 CTGTGCCCTGGGCCTGTAAGCAGGGGAC.........ACCCCCCATGAT. 766

      55 luSerValSerHisValGlyLeuGlnAla............HisPheGlu 67
            :::         ||||||:::::            |||:::
     767 .....CTCCTGGATTGGGGCCTCCGTGTCCTCCCCGGGCCCCACTACTGC 811

      68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
            |||:::  |||       |||           |||:::||
     812 CCGCTCCTCAGTGCTCACCCTTACCCCAAAGCCCCAGGACCACGGCACCA 861

      84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnL 101
         |   :::|||::::::   |||
     862 GCCTCACCTGTCAGGTGACCTT.......................... 883

     101 euLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
                      :::|||:::   |||   |||:::   |||:::
     884 ................GCCTGGGACAGGTGTGACCACGACCAGTACCGT 916

     118 LeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134
            ||||||:::              ::::::|||:::|||:::|||
     917 ...CCGCCTCGA............TGTGTCCTACCCTCCTTGGAACTTGA 951

     134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
```

```
       :::|||
    952 CCATGAC..................................... 958

    151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMet 167
                                    |||           :::
    959 ........................TGTCTTCCAAGGAGATGCCAC 979

    168 SerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGl 184
        |||  |||     ::::::  :::     :::  |||
    980 AGCATCCACAGCCCTGGGAAATGGCTCATCTCTTTCAGTCCTTGAGGGCC 1029

    184 uIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrT 201
         :::::::::  :::|||        |||:::     ||||||
   1030 AGTCTCTGCGCCTGGTCTGTGCTGTCAACAGCAATCCCCCTGCCAGGCTG 1079

    201 rpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu..... 214
          |||:::  ||||||  :::         |||  :::
   1080 AGCTGGACCCGGGGGAGCCTGACCCTGTGCCCCTCACGGTCCTCAAACCC 1129

    215 TyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGl 231
        :::|||:::||||||                ::::||  |||
   1130 TGGGCTGCTGGAGCTGCCTCGAGTGCACGTGAGGGATGAAGGGGAATTCA 1179

    231 yAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
              ||||||  :::::::::     |||
   1180 CCTGCCGAGCTCAGAACGCTCAGGGCTCCCAGCACAT 1216

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-038-832-1

seq_documentation_block:
; Sequence 1, Application US/09038832
; Patent No. 6146845
;  GENERAL INFORMATION:
;    APPLICANT:  KIKLY, KRISTINE
;    APPLICANT:  ERICKSON-MILLER, CONNIE
;    TITLE OF INVENTION:  Sialoadhesin Family Member-2
;    TITLE OF INVENTION:  (SAF-2)
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/038,832
;      FILING DATE:  11-MAR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/041,886
;      FILING DATE:  02-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-50018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2900 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-038-832-1

alignment_scores:
            Quality:   81.50              Length:    229
              Ratio:  0.849                Gaps:      9
 Percent Similarity:  41.921   Percent Identity: 22.271

alignment_block:
 US-09-674-779-2 x US-09-038-832-1   ..

 Align seg 1/1  to: US-09-038-832-1  from: 1  to: 2900

     23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
        |||  |||  :::|||:::|||  |||::::::  |||:::::::||
    537 CCTCATCCTAGGGACCCTAGAGTCTGGCCACTCCAGGAACCTGACCTGCT 586

     39 uIleThrLysAspLys...IleGlyAspHisHisThrHisGluHisAspG 55
        |  :::              ||||||              ||||||
    587 CTGTGCCCTGGGCCTGTAAGCAGGGGAC.....ACCCCCCATGAT. 626

     55 luSerValSerHisValGlyLeuGlnAla.....HisPheGlu 67
          :::         ||||||::::::          |||:::
    627 .....CTCCTGGATTGGGGCCTCCGTGTCCTCCCCGGGCCCCACTACTGC 671

     68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
           |||:::  |||       |||          |||:::||
    672 CCGCTCCTCAGTGCTCACCCTTACCCCAAAGCCCCAGGACCACGGCACCA 721

     84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnL 101
        |  :::|||::::::  |||
    722 GCCTCACCTGTCAGGTGACCTT..... 743

    101 euLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
                        :::|||:::  |||  |||:::  |||:::
    744 ..............GCCTGGGACAGGTGTGACCACGACCAGTACCGT 776

    118 LeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134
          |||||||::                ::::::|||:::|||:::|||
    777 ...CCGCCTCGA...........TGTGTCCTACCCTCCTTGGAACTTGA 811

    134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
         :::|||
    812 CCATGAC..................................... 818

    151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMet 167
                                    |||           :::
    819 ........................TGTCTTCCAAGGAGATGCCAC 839

    168 SerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGl 184
        |||  |||     ::::::  :::     :::  |||
    840 AGCATCCACAGCCCTGGGAAATGGCTCATCTCTTTCAGTCCTTGAGGGCC 889

    184 uIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrT 201
         :::::::::  :::|||        |||:::     ||||||
    890 AGTCTCTGCGCCTGGTCTGTGCTGTCAACAGCAATCCCCCTGCCAGGCTG 939

    201 rpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu..... 214
          |||:::  ||||||  :::         |||  :::
    940 AGCTGGACCCGGGGGAGCCTGACCCTGTGCCCCTCACGGTCCTCAAACCC 989

    215 TyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGl 231
        :::|||:::||||||                ::::||  |||
    990 TGGGCTGCTGGAGCTGCCTCGAGTGCACGTGAGGGATGAAGGGGAATTCA 1039

    231 yAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
              ||||||  :::::::::     |||
   1040 CCTGCCGAGCTCAGAACGCTCAGGGCTCCCAGCACAT 1076

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-439-313-474

seq_documentation_block:
; Sequence 474, Application US/09439313
; Patent No. 6329505
```

```
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 474
;   LENGTH: 1594
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-439-313-474


alignment_scores:
            Quality:   81.00              Length:    209
              Ratio:  0.871                Gaps:     10
 Percent Similarity:  44.498   Percent Identity: 20.096

alignment_block:
 US-09-674-779-2 x US-09-439-313-474   ..

 Align seg 1/1  to: US-09-439-313-474  from: 1  to: 1594

      48 HisHisThrHisGluHisAspGluSerVal......SerHisValGlyLe 62
         ||||||||||||||||||   :::::::::      :::|||
     919 CATCACACACATGAACACACCGACACACTGCCATATGGACACTGGCACAC 968

      62 uGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnG 79
          ::::::|||      ||||||   ::::::|||      |||      :
     969 ACACTGCCACACTGTCACATGGACACACCTCCACACCATCACACCACCAC 1018

      79 luValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeu 95
         ::::::::                 :::::::::|||      :::::
    1019 ACACACTGCCT.............GTGGACACAAGGACACACAGACACTGT 1056

      96 ProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGl 112
                ::::|||      :::|||   |||
    1057 CACACAGATACACAAAACACTGTCACACGGAGA..................... 1089

     112 yHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValP 129
          |||:::      :::   ||||||      ||||||   :::
    1090 .CATCACCATGCAGATACACCACCA......CTCTGGTGCCGT....... 1125

     129 roThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla 145

    1125 ..................................................... 1125

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
                                          ||||||      ||
    1126 .................................CTGAATTACCCTGC 1139

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
         |||||||:::|||:::::               |||:::::::::   |
    1140 TGGGGGGACAGCAGTGGCA............TACTCATGCCTAAGTGACT 1177

     179 rpVal.....................................ProAsp 182
         ||:::                                     |||
    1178 GGCTTTCACCCCAGTAGTGATTGCCCTCCATCAACACTGCCCACCCCAGG 1227

     183 LeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeu..... 197
```

```
         |||   :::      |||   ::::::      |||:::::::::
    1228 TTGGGGCTA...CCCCAGCCCATCTTTACAAAACAGGGCAAGGTGAACTA 1274

     198 .CysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyL 214
          :::   |||   :::      ::::::|||:::   :::   |||
    1275 ATGGAGTGGGTGGAGGAGTTGGAAGAAATCCCAGCGTCAGTCACCGGGAT 1324

     214 euTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrp 230
          :::   |||::::::   |||:::   :::|||||||:::   |||
    1325 AGAATTCCCAAGGAACCCTCTTTTTGGAG...GATGGTTTCCATTTCTGG 1371

     231 GlyAlaGlnPheSerGluThrAsnSer 239
          :::   :::::::::|||:::
    1372 AGGCGATCTGCCGACAGGGTGAATGCC 1398

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-741-881-1

seq_documentation_block:
; Sequence 1, Application US/08741881
; Patent No. 5789245
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Ibanez, Carlos E.
;    APPLICANT:  Chang, Stephen M.W.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    APPLICANT:  Belli, Barbara A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/741,881
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  930049.423C6 / 1146.007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16656 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-741-881-1


alignment_scores:
            Quality:   81.00              Length:    210
              Ratio:  0.853                Gaps:     14
 Percent Similarity:  45.238   Percent Identity: 25.238

alignment_block:
 US-09-674-779-2 x US-08-741-881-1   ..

 Align seg 1/1  to: US-08-741-881-1  from: 1  to: 16656

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
```

```
      |||||||   |||       |||   :::|||||||              |||::
   8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698

     27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
        :|||       |||||:::|||       :::|||             |||:
   8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 8739

     44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
        ::       |||  ||||||||||||          :::   |||
   8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 8782

     60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                |||::::::      :::::::::
   8783 .........................TGGATACGACCAAAGCGGAGCAGC 8806

     77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
        |||::::::   ||| ||||||::::::      |||
   8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTA 8856

     88 ................................................... 88

   8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

     89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
              |||       |||   :::|||             ||||||
   8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

     98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
          ::::::      ::::::::::::::|||      :::|||
   8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

    112 ............................GlyHisGluProTyrGlnLeu 118
                                    |||:::|||   |||:::|||
   9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

    119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
        ||||||   |||       |||:::   :::|||          :::|||::
   9057 CCTCCCGTTCAC......GGTAAAAGAATTCCTTGCACA...GTGTACGA 9097

    135 nAspLeuLys...SerArgGlyIleLeu............ProAla.Asn 146
        :   ||||||   :::   |||   :::              |||::: :::
   9098 CCGTCTGAAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGC 9147

    147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
        |||   |||:::   ::::::::::|||   :::                ||
   9148 ACGCTTATACATCCTACCTGGAAGAATCATCAG..............GG 9182

    163 yGlyAlaAlaMetSerLysHisLeu 171
        |      ::::::|||:::||||||
   9183 AAAGTTTACGCAAAGCCGCCATCTG 9207

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-739-158-1

seq_documentation_block:
; Sequence 1, Application US/08739158
; Patent No. 5814482
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,158
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  930049.423D3 / 1146.012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16656 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-739-158-1


alignment_scores:
            Quality:   81.00              Length:    210
              Ratio:  0.853                Gaps:     14
 Percent Similarity:  45.238   Percent Identity: 25.238

alignment_block:
 US-09-674-779-2 x US-08-739-158-1   ..

 Align seg 1/1  to: US-08-739-158-1  from: 1  to: 16656

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||       |||   :::||||||             |||::
    8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
         :|||       |||||:::|||       :::|||             |||:
    8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 8739

      44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
         ::       |||  ||||||||||||          :::   |||
    8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 8782

      60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                 |||::::::      :::::::::
    8783 .........................TGGATACGACCAAAGCGGAGCAGC 8806

      77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
         |||::::::   ||| ||||||::::::      |||
    8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTA 8856

      88 ................................................... 88

    8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

      89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
               |||       |||   :::|||             ||||||
    8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

      98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
           ::::::      ::::::::::::::|||      :::|||
    8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

     112 ............................GlyHisGluProTyrGlnLeu 118
                                     |||:::|||   |||:::|||
    9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
         ||||||   |||       |||:::   :::|||          :::|||::
    9057 CCTCCCGTTCAC......GGTAAAAGAATTCCTTGCACA...GTGTACGA 9097

     135 nAspLeuLys...SerArgGlyIleLeu............ProAla.Asn 146
```

```
       :   ||||||   :::   |||   :::            |||::: :::
    9098 CCGTCTGAAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGC 9147

     147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
         |||   |||:::   ::::::::::|||   :::              ||
    9148 ACGCTTATACATCCTACCTGGAAGAATCATCAG.............GG 9182

     163 yGlyAlaAlaMetSerLysHisLeu 171
         |     :::::::|||:::||||||
    9183 AAAGTTTACGCAAAGCCGCCATCTG 9207

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-739-167-1

seq_documentation_block:
; Sequence 1, Application US/08739167
; Patent No. 5843723
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Ibanez, Carlos E.
;    APPLICANT:  Chang, Stephen M.W.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    APPLICANT:  Belli, Barbara A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS  AND ALPHAVIRUS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,167
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  930049.423C7 / 1146.008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16656 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-739-167-1


alignment_scores:
            Quality:   81.00            Length:    210
              Ratio:  0.853              Gaps:     14
 Percent Similarity:  45.238   Percent Identity: 25.238

alignment_block:
 US-09-674-779-2 x US-08-739-167-1   ..

 Align seg 1/1  to: US-08-739-167-1  from: 1  to: 16656

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||      |||   :::||||||           |||::
    8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44

         :|||     ||||||:::|||      :::|||           |||:
    8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.....GACG 8739

      44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
         ::      |||  ||||||||||||         :::  |||
    8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT..... 8782

      60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                  |||::::::     :::::::::
    8783 .........................TGGATACGACCAAAGCGGAGCAGC 8806

      77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln..... 88
         |||::::::   ||| ||||||::::::      |||
    8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTA 8856

      88 ..... 88

    8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

      89 ......SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
               |||      |||   :::|||          ||||||
    8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

      98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
           ::::::      ::::::::::::::::|||      :::|||
    8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

     112 ............................GlyHisGluProTyrGlnLeu 118
                                     |||:::|||   |||:::|||
    9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
         ||||||   |||      |||:::   :::|||          :::|||::
    9057 CCTCCCGTTCAC......GGTAAAAGAATTCCTTGCACA...GTGTACGA 9097

     135 nAspLeuLys...SerArgGlyIleLeu.....ProAla.Asn 146
         :   ||||||   :::   |||   :::            |||::: :::
    9098 CCGTCTGAAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGC 9147

     147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
         |||   |||:::   ::::::::::|||   :::              ||
    9148 ACGCTTATACATCCTACCTGGAAGAATCATCAG.............GG 9182

     163 yGlyAlaAlaMetSerLysHisLeu 171
         |     :::::::|||:::||||||
    9183 AAAGTTTACGCAAAGCCGCCATCTG 9207

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-404-796-1

seq_documentation_block:
; Sequence 1, Application US/08404796
; Patent No. 6015686
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Ibanez, Carlos E.
;    APPLICANT:  Chang, Stephen M.W.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    APPLICANT:  Belli, Barbara A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/404,796
;       FILING DATE:  15-MAR-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McMasters, David D.
;       REGISTRATION NUMBER:  33,963
;       REFERENCE/DOCKET NUMBER:  930049.423C5 / 1146.006
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  16656 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-404-796-1


alignment_scores:
            Quality:   81.00              Length:    210
              Ratio:  0.853                Gaps:     14
 Percent Similarity:  45.238   Percent Identity: 25.238

alignment_block:
 US-09-674-779-2 x US-08-404-796-1   ..

 Align seg 1/1  to: US-08-404-796-1  from: 1  to: 16656

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||      |||   :::||||||            |||::
    8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
         :|||      ||||||:::|||      :::|||          |||:
    8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 8739

      44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
         ::      |||  ||||||||||||        :::   |||
    8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 8782

      60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                |||::::::      :::::::::
    8783 ..........................TGGATACGACCAAAGCGGAGCAGC 8806

      77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
         |||::::::   ||| ||||||::::::      |||
    8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTA 8856

      88 ................................................. 88

    8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

      89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
               |||      |||   :::|||            ||||||
    8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

      98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
             ::::::      :::::::::::::::|||      :::|||
    8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

     112 ..............................GlyHisGluProTyrGlnLeu 118
                                     |||:::|||   |||:::|||
    9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
         ||||||   |||      |||:::   :::|||         :::|||::
    9057 CCTCCCGTTCAC......GGTAAAAGAATTCCTTGCACA...GTGTACGA 9097

     135 nAspLeuLys...SerArgGlyIleLeu............ProAla.Asn 146
         :   ||||||   :::   |||   :::            |||::: :::

    9098 CCGTCTGAAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGC 9147

     147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
         |||   |||:::   :::::::::|||   :::              ||
    9148 ACGCTTATACATCCTACCTGGAAGAATCATCAG.............GG 9182

     163 yGlyAlaAlaMetSerLysHisLeu 171
         |      ::::::|||:::|||||
    9183 AAAGTTTACGCAAAGCCGCCATCTG 9207

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-931-869-1

seq_documentation_block:
; Sequence 1, Application US/08931869
; Patent No. 6015694
;  GENERAL INFORMATION:
;    APPLICANT:  Dubensky Jr, Thomas W
;    APPLICANT:  Polo, John M.
;    APPLICANT:  Ibanez, Carlos E.
;    APPLICANT:  Chang, Stephen M.W.
;    APPLICANT:  Jolly, Douglas J.
;    APPLICANT:  Driver, David A.
;    APPLICANT:  Belli, Barbara A.
;    TITLE OF INVENTION:  EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;    NUMBER OF SEQUENCES:  128
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,869
;      FILING DATE:  16-SEP-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/404,796
;      FILING DATE:  15-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  930049.423C5 / 1146.006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16656 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-931-869-1


alignment_scores:
            Quality:   81.00              Length:    210
              Ratio:  0.853                Gaps:     14
 Percent Similarity:  45.238   Percent Identity: 25.238

alignment_block:
 US-09-674-779-2 x US-08-931-869-1   ..

 Align seg 1/1  to: US-08-931-869-1  from: 1  to: 16656

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||      |||   :::||||||            |||::
    8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698
```

```
     27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
        :|||      ||||||:::|||      :::|||            |||:
   8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 8739

     44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
        ::      |||  ||||||||||||         :::   |||
   8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 8782

     60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                 |||::::::      :::::::::
   8783 .......................TGGATACGACCAAAGCGGAGCAGC 8806

     77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
        |||::::::   ||| ||||||::::::      |||
   8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTA 8856

     88 ..................................................... 88

   8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

     89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
             |||      |||   :::|||            ||||||
   8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

     98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
         ::::::      ::::::::::::::::|||      :::|||
   8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

    112 ............................GlyHisGluProTyrGlnLeu 118
                                    |||:::|||   |||:::|||
   9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

    119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
        ||||||   |||      |||:::   :::|||         :::|||::
   9057 CCTCCCGTTCAC......GGTAAAAGAATTCCTTGCACA...GTGTACGA 9097

    135 nAspLeuLys...SerArgGlyIleLeu............ProAla.Asn 146
        :   ||||||   :::   |||   :::            |||::: :::
   9098 CCGTCTGAAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGC 9147

    147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
        |||   |||:::   :::::::::|||   :::              ||
   9148 ACGCTTATACATCCTACCTGGAAGAATCATCAG..............GG 9182

    163 yGlyAlaAlaMetSerLysHisLeu 171
        |      ::::::|||:::||||||
   9183 AAAGTTTACGCAAAGCCGCCATCTG 9207

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-350-399-1

seq_documentation_block:
; Sequence 1, Application US/09350399
; Patent No. 6342372
;   GENERAL INFORMATION:
;        APPLICANT: Dubensky Jr, Thomas W
;                   Polo, John M.
;                   Jolly, Douglas J.
;                   Driver, David A.
;        TITLE OF INVENTION: EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
;        NUMBER OF SEQUENCES: 128
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: SEED and BERRY LLP
;             STREET: 6300 Columbia Center, 701 Fifth Avenue
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: US
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/350,399
;             FILING DATE: 08-Jul-1999
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McMasters, David D.
;             REGISTRATION NUMBER: 33,963
;             REFERENCE/DOCKET NUMBER: 930049.423D1 / 1146.010
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 16656 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-350-399-1


alignment_scores:
            Quality:   81.00              Length:    210
              Ratio:  0.853                Gaps:     14
 Percent Similarity:  45.238   Percent Identity: 25.238

alignment_block:
 US-09-674-779-2 x US-09-350-399-1   ..

 Align seg 1/1  to: US-09-350-399-1  from: 1  to: 16656

     11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
        ||||||   |||      |||   :::|||||||            |||::
   8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698

     27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
        :|||      ||||||:::|||      :::|||            |||:
   8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 8739

     44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
        ::      |||  ||||||||||||         :::   |||
   8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 8782

     60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                 |||::::::      :::::::::
   8783 .......................TGGATACGACCAAAGCGGAGCAGC 8806

     77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
        |||::::::   ||| ||||||::::::      |||
   8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTA 8856

     88 ..................................................... 88

   8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

     89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
             |||      |||   :::|||            ||||||
   8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

     98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
         ::::::      ::::::::::::::::|||      :::|||
   8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

    112 ............................GlyHisGluProTyrGlnLeu 118
                                    |||:::|||   |||:::|||
   9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

    119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
        ||||||   |||      |||:::   :::|||         :::|||::
   9057 CCTCCCGTTCAC......GGTAAAAGAATTCCTTGCACA...GTGTACGA 9097

    135 nAspLeuLys...SerArgGlyIleLeu............ProAla.Asn 146
        :   ||||||   :::   |||   :::            |||::: :::
```

```
    9098 CCGTCTGAAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGACCGC 9147

     147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
         |||   |||:::   ::::::::::|||   :::             ||
    9148 ACGCTTATACATCCTACCTGGAAGAATCATCAG.............GG 9182

     163 yGlyAlaAlaMetSerLysHisLeu 171
         |      ::::::|||:::|||||
    9183 AAAGTTTACGCAAAGCCGCCATCTG 9207

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-276-452A-66

seq_documentation_block:
; Sequence 66, Application US/08276452A
; Patent No. 5646029
;  GENERAL INFORMATION:
;    APPLICANT:  Chen, Chao-Guang
;    APPLICANT:  Mau, Shaio-Lim
;    APPLICANT:  Du, He
;    APPLICANT:  Gane, Alison M
;    APPLICANT:  Bacic, Antony
;    APPLICANT:  Clarke, Adrienne E
;    TITLE OF INVENTION:  Plant Arabinogalactan Protein (AGP) Genes
;    NUMBER OF SEQUENCES:  91
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  United States of America
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,452A
;      FILING DATE:  18-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Caruthers, Jennie M.
;      REGISTRATION NUMBER:  34,464
;      REFERENCE/DOCKET NUMBER:  27-91A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303)499-8080
;      TELEFAX:  (303)499-8089
;      TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  893 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..504
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  70..138
;      OTHER INFORMATION:  /note= "Putative secretion signal
;      OTHER INFORMATION:  peptide"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  70..339
;      OTHER INFORMATION:  /note= "Amino acids 70-138, 38-53,
;      OTHER INFORMATION:  and 71-90 are sequences which match the peptide
;      OTHER INFORMATION:  sequences obtained by protein sequencing"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  87..324
;      OTHER INFORMATION:  /note= "Amino acids 40, 44, 45, 49,
;      OTHER INFORMATION:  50, 74, 76, 80, 81, and 85 are hydroxylated
;      OTHER INFORMATION:  prolines"
US-08-276-452A-66


alignment_scores:
            Quality:   80.50              Length:    197
              Ratio:  0.982                Gaps:     11
 Percent Similarity:  41.624   Percent Identity: 24.365

alignment_block:
 US-09-674-779-2 x US-08-276-452A-66   ..

 Align seg 1/1  to: US-08-276-452A-66  from: 1  to: 893

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
         |||   |||   :::   |||       |||   :::   |||
     170 CCCCAACCGCAAAGTCGCCAACCGCCACCCCACCAACCGCCA........ 211

      39 uIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluS 56
                                          |||:::|||:::
     212 ................................CACCGCCATCAGCCG. 226

      56 erValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMet 72
                                                  |||:::
     227 ........................................TACCAGTTC 235

      73 HisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSe 89
         ||||||   :::|||::::::      |||:::|||   :::   |||
     236 CATCACCCAGCAAAACACCAA...CCGCGTCACCAACTCCATCACCAGTG 282

      89 rArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaA 106
          :::
     283 ACAG..................................................... 286

     106 rgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHis 122
                              |||:::|||   |||:::|||||||:::|||
     287 ...................CACCAACCC...CAAGTGCCTCCCCACCAT 313

     123 LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGln...AspLeuLy 138
         |||      |||:::   |||   |||   |||   |||   :::|||
     314 CTTCCACACCAGCTTCCACCCCAGCTTCCACTCCAGCAGCTAAGTCTCCA 363

     138 sSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgA 155
             |||   :::||||||   :::::::|||   |||:::   :::|||
     364 TCGTCGTCAGCTGCTCCCT...CAGGCTCAAGCCCGAACTCCCCACCGGC 410

     155 snProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeu 171
                 :::   |||   |||         :::   :::      |||
     411 TGACGCTATTCCTCCAAGTGGCACCTCCGCCATCAGCCGCGTTGCTATTG 460

     172 ThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGl 188
                                        ||||||:::
     461 ............................CTGGAACTG........ 469

     188 nAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisG 205
             |||   |||:::::::::|||||||||      ||||||:::
     470 .CTCTTGCTGGAGTTTTCTTCGCGATTGTGT......TGGCTTAGATTC. 511

     205 lyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGly 218
                 ::::::|||:::|||:::   :::       |||
     512 ........ATGGGATTTGCTCTTTCGGGTTTTCCTATTGGT 544

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-798-744-66

seq_documentation_block:
; Sequence 66, Application US/08798744
; Patent No. 5830747
;  GENERAL INFORMATION:
;    APPLICANT:  Chen, Chao-Guang
;    APPLICANT:  Mau, Shaio-Lim
;    APPLICANT:  Du, He
;    APPLICANT:  Gane, Alison M
```

```
;    APPLICANT:  Bacic, Antony
;    APPLICANT:  Clarke, Adrienne E
;    TITLE OF INVENTION:  Plant Arabinogalactan Protein (AGP) Genes
;    NUMBER OF SEQUENCES:  91
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  United States of America
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/798,744
;      FILING DATE:  13-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/276,452
;      FILING DATE:  18-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Caruthers, Jennie M.
;      REGISTRATION NUMBER:  34,464
;      REFERENCE/DOCKET NUMBER:  27-91A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303)499-8080
;      TELEFAX:  (303)499-8089
;      TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  893 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..504
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  70..138
;      OTHER INFORMATION:  /note= "Putative secretion signal
;      OTHER INFORMATION:  peptide"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  70..339
;      OTHER INFORMATION:  /note= "Amino acids 70-138, 38-53,
;      OTHER INFORMATION:  and 71-90 are sequences which match the peptide
;      OTHER INFORMATION:  sequences obtained by protein sequencing"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  87..324
;      OTHER INFORMATION:  /note= "Amino acids 40, 44, 45, 49,
;      OTHER INFORMATION:  50, 74, 76, 80, 81, and 85 are hydroxylated
;      OTHER INFORMATION:  prolines"
US-08-798-744-66


alignment_scores:
            Quality:   80.50              Length:    197
              Ratio:  0.982                Gaps:     11
 Percent Similarity:  41.624   Percent Identity: 24.365

alignment_block:
 US-09-674-779-2 x US-08-798-744-66   ..

 Align seg 1/1  to: US-08-798-744-66  from: 1  to: 893

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
         |||   |||   :::   |||       |||   :::   |||

     170 CCCCAACCGCAAAGTCGCCAACCGCCACCCCACCAACCGCCA..... 211

      39 uIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluS 56
                                                |||:::|||:::
     212 ..............................CACCGCCATCAGCCG. 226

      56 erValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMet 72
                                                      |||:::
     227 ........................................TACCAGTTC 235

      73 HisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSe 89
         ||||||   :::|||::::::       |||:::|||   :::   |||
     236 CATCACCCAGCAAAACACCAA...CCGCGTCACCAACTCCATCACCAGTG 282

      89 rArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaA 106
          :::
     283 ACAG..................................................... 286

     106 rgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHis 122
                              |||:::|||   |||:::||||||:::|||
     287 ....................CACCAACCC...CAAGTGCCTCCCCACCAT 313

     123 LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGln...AspLeuLy 138
         |||       |||:::   |||   |||   |||   |||   :::|||
     314 CTTCCACACCAGCTTCCACCCCAGCTTCCACTCCAGCAGCTAAGTCTCCA 363

     138 sSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgA 155
             |||   :::||||||   ::::::|||   |||:::   :::|||
     364 TCGTCGTCAGCTGCTCCCT...CAGGCTCAAGCCCGAACTCCCCACCGGC 410

     155 snProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeu 171
                 :::   |||   |||         :::   :::       |||
     411 TGACGCTATTCCTCCAAGTGGCACCTCCGCCATCAGCCGCGTTGCTATTG 460

     172 ThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGl 188
                                           ||||||:::
     461 ..............................CTGGAACTG........ 469

     188 nAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisG 205
             |||   |||:::::::::|||||||||       ||||||:::
     470 .CTCTTGCTGGAGTTTTCTTCGCGATTGTGT......TGGCTTAGATTC. 511

     205 lyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGly 218
                 :::::::|||:::|||:::   :::       |||
     512 ........ATGGGATTTGCTCTTTCGGGTTTTCCTATTGGT 544

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-203-532F-3

seq_documentation_block:
; Sequence 3, Application US/08203532F
; Patent No. 5856121
;  GENERAL INFORMATION:
;    APPLICANT:  Gorski, David H.
;    APPLICANT:  Walsh, Kenneth
;    TITLE OF INVENTION:  Growth Arrest Homeobox Gene
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter, and Griswold
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/203,532F
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Golrick, Mary E.
;       REGISTRATION NUMBER:  34829
;       REFERENCE/DOCKET NUMBER:  22311/00114
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (216) 622-8200
;       TELEFAX:  (216) 241-0816
;       TELEX:  980499
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  941 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  33..941
US-08-203-532F-3


alignment_scores:
            Quality:   80.50              Length:    180
              Ratio:  1.019                Gaps:     10
 Percent Similarity:  43.889   Percent Identity: 21.667

alignment_block:
 US-09-674-779-2 x US-08-203-532F-3   ..

 Align seg 1/1  to: US-08-203-532F-3  from: 1  to: 941

      48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
         ||||||   |||:::|||   ::::::   :::|||   :::|||||||:: 
     246 CACCACCACCATCACCACCATCAGCAGCAGCAGCACCAGGCTCTGCAAAC 295

      64 aHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValV 81
         ::::         |||         :::   ::::::       ::::::: 
     296 CAAC.........TGGCACCTCCCGCAGATGTCTTCCCCACCGAGTGCGG 336

      81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
         ::|||::::::                                |||||||||
     337 CTCGGCATAGCCT............................CTGCCTCCA 358

      98 MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAla......CysGl 112
                               ||||||   |||:::|||         ||
     359 ......................GCCCGACTCTGGAGGGCCCCCAGAGTTGG 387

     112 yHisGluPro.....................TyrGlnLeuProProGluH 122
         |::::::|||                    :::||||||       :::| 
     388 GGAGCAGCCCGCCCGTCCTGTGCTCCAACTCTTCCAGCTTGGGCTCCAGC 437

     122 isLeu...TrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu 137
         ||      ||||||::::::                              
     438 ACCCCGACTGGGGCCGCGTG............................... 457

     138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
         ::::::   |||:::   |||   :::|||                     
     458 CGCGCCGGGGGACTACGGCCGCCAGGCACTGTC................... 490

     154 gAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
                           |||:::||||||   |||      ::::::   
     491 ................ACCTGCGGAGGCGGAGAAGCGAAGCGGCGGCAAG 524

     171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
             :::      :::   :::                |||::::::::|||
     525 AGGAAAAGCGACAGCTCAGACTCCCAGGAAGGAAATTACAAGTCAGAAGT 574

     188 Gln.......................AlaLeuTyrGluLeuGlnAsn.. 195
         |||                      :::::::|||:::      |||   
     575 CAACAGCAAACCCAGGAAAGAAAGGACAGCATTTACCAAAGAGCAAATCA 624

     196 .............ArgLeuCysGlnTyrTrpLeuGluHis 204
                                 |||:::|||       |||:::|||
     625 GAGAACTTGAAGCAGAATTTGCCCATCATAATTATCTCAC 664

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-078-465-3

seq_documentation_block:
; Sequence 3, Application US/09078465
; Patent No. 6280969
;  GENERAL INFORMATION:
;    APPLICANT:  Gorski, David H.
;    APPLICANT:  Walsh, Kenneth
;    TITLE OF INVENTION:  Growth Arrest Homeobox Gene
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter, and Griswold
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,465
;      FILING DATE:  14-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/203,532
;      FILING DATE:  24-FEB-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22311/00114
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 622-8200
;      TELEFAX:  (216) 241-0816
;      TELEX:  980499
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  941 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..941
US-09-078-465-3


alignment_scores:
            Quality:   80.50              Length:    180
              Ratio:  1.019                Gaps:     10
 Percent Similarity:  43.889   Percent Identity: 21.667

alignment_block:
 US-09-674-779-2 x US-09-078-465-3   ..

 Align seg 1/1  to: US-09-078-465-3  from: 1  to: 941

      48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
         ||||||   |||:::|||   ::::::   :::|||   :::|||||||:: 
     246 CACCACCACCATCACCACCATCAGCAGCAGCAGCACCAGGCTCTGCAAAC 295

      64 aHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValV 81
         ::::         |||         :::   ::::::       ::::::: 
     296 CAAC.........TGGCACCTCCCGCAGATGTCTTCCCCACCGAGTGCGG 336
```

```
     81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
        ::|||::::::                                 |||||||||
    337 CTCGGCATAGCCT.........................CTGCCTCCA 358

     98 MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAla......CysGl 112
                          ||||||   |||:::|||         ||
    359 ....................GCCCGACTCTGGAGGGCCCCCAGAGTTGG 387

    112 yHisGluPro.....................TyrGlnLeuProProGluH 122
        |:::::::|||                  :::|||||||      :::|
    388 GGAGCAGCCCGCCCGTCCTGTGCTCCAACTCTTCCAGCTTGGGCTCCAGC 437

    122 isLeu...TrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu 137
        ||      ||||||::::::
    438 ACCCCGACTGGGGCCGCGTG............................. 457

    138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
        ::::::   |||:::   |||   :::|||
    458 CGCGCCGGGGGACTACGGCCGCCAGGCACTGTC................. 490

    154 gAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
                         |||:::||||||   |||      ::::::
    491 ................ACCTGCGGAGGCGGAGAAGCGAAGCGGCGGCAAG 524

    171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
            :::      :::   :::            |||:::::::::|||
    525 AGGAAAAGCGACAGCTCAGACTCCCAGGAAGGAAATTACAAGTCAGAAGT 574

    188 Gln.......................AlaLeuTyrGluLeuGlnAsn.. 195
        |||                      ::::::|||:::      |||
    575 CAACAGCAAACCCAGGAAAGAAAGGACAGCATTTACCAAAGAGCAAATCA 624

    196 .............ArgLeuCysGlnTyrTrpLeuGluHis 204
                    |||:::|||         |||:::|||
    625 GAGAACTTGAAGCAGAATTTGCCCATCATAATTATCTCAC 664

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-01882A-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9501882A
;  GENERAL INFORMATION:
;    APPLICANT:  Gorski, David H.
;    APPLICANT:  Walsh, Kenneth
;    TITLE OF INVENTION:  Growth Arrest Homeobox Gene
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter, and Griswold
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/01882A
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22311/00114
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 622-8200
;      TELEFAX:  (216) 241-0816
;      TELEX:  980499
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  941 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..941
PCT-US95-01882A-3


alignment_scores:
           Quality:   80.50            Length:    180
             Ratio:  1.019              Gaps:     10
 Percent Similarity:  43.889   Percent Identity: 21.667

alignment_block:
 US-09-674-779-2 x PCT-US95-01882A-3   ..

 Align seg 1/1  to: PCT-US95-01882A-3  from: 1  to: 941

     48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
        ||||||   |||:::|||   ::::::   :::|||   :::|||||::
    246 CACCACCACCATCACCACCATCAGCAGCAGCAGCACCAGGCTCTGCAAAC 295

     64 aHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValV 81
        ::::         |||        :::   ::::::      :::::::
    296 CAAC.........TGGCACCTCCCGCAGATGTCTTCCCCACCGAGTGCGG 336

     81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
        ::|||::::::                                 |||||||||
    337 CTCGGCATAGCCT.........................CTGCCTCCA 358

     98 MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAla......CysGl 112
                          ||||||   |||:::|||         ||
    359 ....................GCCCGACTCTGGAGGGCCCCCAGAGTTGG 387

    112 yHisGluPro.....................TyrGlnLeuProProGluH 122
        |:::::::|||                  :::|||||||      :::|
    388 GGAGCAGCCCGCCCGTCCTGTGCTCCAACTCTTCCAGCTTGGGCTCCAGC 437

    122 isLeu...TrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu 137
        ||      ||||||::::::
    438 ACCCCGACTGGGGCCGCGTG............................. 457

    138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
        ::::::   |||:::   |||   :::|||
    458 CGCGCCGGGGGACTACGGCCGCCAGGCACTGTC................. 490

    154 gAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
                         |||:::||||||   |||      ::::::
    491 ................ACCTGCGGAGGCGGAGAAGCGAAGCGGCGGCAAG 524

    171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
            :::      :::   :::            |||:::::::::|||
    525 AGGAAAAGCGACAGCTCAGACTCCCAGGAAGGAAATTACAAGTCAGAAGT 574

    188 Gln.......................AlaLeuTyrGluLeuGlnAsn.. 195
        |||                      ::::::|||:::      |||
    575 CAACAGCAAACCCAGGAAAGAAAGGACAGCATTTACCAAAGAGCAAATCA 624

    196 .............ArgLeuCysGlnTyrTrpLeuGluHis 204
                    |||:::|||         |||:::|||
    625 GAGAACTTGAAGCAGAATTTGCCCATCATAATTATCTCAC 664

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-524-2

seq_documentation_block:
; Sequence 2, Application US/09479524
; Patent No. 6268350
; GENERAL INFORMATION:
;  APPLICANT: Barbera-Guillem, Emilio
;  TITLE OF INVENTION: Polynucleotides For Inhibiting Metastasis And Tumor Cell Growt
```

```
;  FILE REFERENCE: B-37
;  CURRENT APPLICATION NUMBER: US/09/479,524
;  CURRENT FILING DATE: 2000-01-07
;  EARLIER APPLICATION NUMBER: US 09/170,948
;  EARLIER FILING DATE: 1998-10-13
;  EARLIER APPLICATION NUMBER: US 60/062,733
;  EARLIER FILING DATE: 1997-10-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: WordPerfect
; SEQ ID NO 2
;   LENGTH: 940
;   TYPE: DNA
;   ORGANISM: murine
US-09-479-524-2


alignment_scores:
           Quality:   80.00            Length:    253
             Ratio:  0.816              Gaps:     16
 Percent Similarity:  38.735   Percent Identity: 25.296

alignment_block:
 US-09-674-779-2 x US-09-479-524-2   ..

 Align seg 1/1  to: US-09-479-524-2  from: 1  to: 940

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSer..... 37
         |||   |||      |||   |||||||||:::      ||||||
     171 CCGCCACCTCCACCACCACCTGTGTCACCACTACCACCGCCATCACAACC 220

      38 ..........ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
                      :::|||      |||      ||||||:::|||:
     221 ACTCCCACTGCCGCCACTGACCCCTCTAAAGAAGAAGGACCACAACACAA 270

      51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
         ::
     271 AT..................................................... 272

      68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
            ||||||                          ||||||   :::::
     273 CTGTGGCTA.....................CCGGTGGTATTTTTCAT 298

      84 n.........AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         :         |||  :::         ||||||   |||
     299 GGTTCTGGTGGCTCTGGTTGGAATGGGATTAGGAATGTAT.......... 338

      98 etSerGlnLeuLeu.ThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
            ||||||::: |||      |||:::||||||           ::
     339 .....CAGCTCTTCCACCTGCAGAAGGAACTGGCAGAACTCCGTGAGTTC 383

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
         :|||   :::           ||||||      :::      ||||||
     384 ACCAACCAAAGCCTTAAAGTATCATCTTTTGAAAAGCAAATAGCCAACCC 433

     131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
           :::|||      :::      ||||||||||||:::   |||    ||
     434 CAGTACACCCTCTGAAAAAA...AAGAGCCGAGGAGTGTGGCCCATTTAA 480

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
         |                    |||:::|||   |||:::         :
     481 C.....................AGGGAACCCCCACTCAAGGTCCATCCCT 509

     163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
         ::|||      |||::::::|||:::   |||   :::   |||:::|||
     510 CTGGA......ATGGGAAGACACATATGGAACCGCTCT...GATCTCTGG 550

     179 ..................................................... 179

     551 AGTGAAGTATAAGAAAGGTGGCCTTGTGATCAACGAAACTGGGTTGTACT 600

     180 ..................ValProAspLeuGluIle............. 185
                            :::|||         :::
     601 TCGTGTATTCCAAAGTATACTTCCGGGGTCAGTCTTGCAACAACCAGCCC 650

     186 ..LysSerGlnAlaLeuTyrGlu........................... 192
           |||   |||:::|||||||||
     651 CTAAACCACAAGGTCTATATGAGGAACTCTAAGTATCCTGAGGATCTGGT 700

     193 ............LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGl 205
                        :::   :::|||   :::||||||   :::   ||
     701 GCTAATGGAGGAGAAGAGGTTGAACTACTGCACTACTGGCCAGATATGGG 750

     205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
         |   :::|||    :::|||      :::      |||:::|||   |||
     751 CCCACAGCAGCTACCTGGGGGCAGTATTCAATCTTACCAGTGCTGACCAT 800

     222 Leu 222
         |||
     801 TTA 803

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-524-8

seq_documentation_block:
; Sequence 8, Application US/09479524
; Patent No. 6268350
; GENERAL INFORMATION:
;  APPLICANT: Barbera-Guillem, Emilio
;  TITLE OF INVENTION: Polynucleotides For Inhibiting Metastasis And Tumor Cell Growt
;  FILE REFERENCE: B-37
;  CURRENT APPLICATION NUMBER: US/09/479,524
;  CURRENT FILING DATE: 2000-01-07
;  EARLIER APPLICATION NUMBER: US 09/170,948
;  EARLIER FILING DATE: 1998-10-13
;  EARLIER APPLICATION NUMBER: US 60/062,733
;  EARLIER FILING DATE: 1997-10-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: WordPerfect
; SEQ ID NO 8
;   LENGTH: 940
;   TYPE: DNA
;   ORGANISM: murine
US-09-479-524-8


alignment_scores:
           Quality:   80.00            Length:    253
             Ratio:  0.816              Gaps:     16
 Percent Similarity:  38.735   Percent Identity: 25.296

alignment_block:
 US-09-674-779-2 x US-09-479-524-8/rev   ..

 Align seg 1/1  to reverse of: US-09-479-524-8  from: 1  to: 940

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSer..... 37
         |||   |||      |||   |||||||||:::      ||||||
     770 CCGCCACCTCCACCACCACCTGTGTCACCACTACCACCGCCATCACAACC 721

      38 ..........ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
                      :::|||      |||      ||||||:::|||:
     720 ACTCCCACTGCCGCCACTGACCCCTCTAAAGAAGAAGGACCACAACACAA 671

      51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
         ::
     670 AT..................................................... 669

      68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
            ||||||                          ||||||   :::::
     668 CTGTGGCTA.....................CCGGTGGTATTTTTCAT 643

      84 n.........AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         :         |||  :::         ||||||   |||
     642 GGTTCTGGTGGCTCTGGTTGGAATGGGATTAGGAATGTAT.......... 603

      98 etSerGlnLeuLeu.ThrThrAlaArgSerTrpGlnAlaCysGlyHisGl 114
            ||||||::: |||      |||:::||||||           ::
```

```
     602 .....CAGCTCTTCCACCTGCAGAAGGAACTGGCAGAACTCCGTGAGTTC 558

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
     557 ACCAACCAAAGCCTTAAAGTATCATCTTTTGAAAAGCAAATAGCCAACCC 508

     131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
     507 CAGTACACCCTCTGAAAAAA...AAGAGCCGAGGAGTGTGGCCCATTTAA 461

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
     460 C.....AGGGAACCCCCACTCAAGGTCCATCCCT 432

     163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
     431 CTGGA......ATGGGAAGACACATATGGAACCGCTCT...GATCTCTGG 391

     179 ..... 179
     390 AGTGAAGTATAAGAAAGGTGGCCTTGTGATCAACGAAACTGGGTTGTACT 341

     180 .....ValProAspLeuGluIle..... 185
     340 TCGTGTATTCCAAAGTATACTTCCGGGGTCAGTCTTGCAACAACCAGCCC 291

     186 ..LysSerGlnAlaLeuTyrGlu..... 192
     290 CTAAACCACAAGGTCTATATGAGGAACTCTAAGTATCCTGAGGATCTGGT 241

     193 .....LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGl 205
     240 GCTAATGGAGGAGAAGAGGTTGAACTACTGCACTACTGGCCAGATATGGG 191

     205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
     190 CCCACAGCAGCTACCTGGGGGCAGTATTCAATCTTACCAGTGCTGACCAT 141

     222 Leu 222
     140 TTA 138

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-846-762-1

seq_documentation_block:
; Sequence 1, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
;  APPLICANT: Lam, Joseph S.
;  APPLICANT: Burrows, Lori
;  APPLICANT: Charter, Deborah
;  APPLICANT: de Kievit, Teresa
;  TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
;  TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
;  FILE REFERENCE: 6580-089
;  CURRENT APPLICATION NUMBER: US/08/846,762A
;  CURRENT FILING DATE: 1997-04-30
;  NUMBER OF SEQ ID NOS: 100
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 24417
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-08-846-762-1


alignment_scores:
            Quality:   80.00              Length:    364
              Ratio:  0.526                Gaps:     19
 Percent Similarity:  41.758   Percent Identity: 19.505

alignment_block:
 US-09-674-779-2 x US-08-846-762-1/rev   ..
```

```
Align seg 1/1  to reverse of: US-08-846-762-1  from: 1  to: 24417

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
    5860 TTGCGCTCGATCAACGAGCGGGGGTTGTAAACGTTGGTAAATACCATGCT 5811

      17 uValAlaCysSerAlaPro..... 23
    5810 CGGCCCGCAGAACACGCCCTCTTCGAGAGTGACATTGTCATATACCGACA 5761

      24 .....IleProThr 26
    5760 CGTTGTTCTGGATCTTGCAGCGATCACCAATAACGACCTTGTTGCCGACG 5711

      27 .....AsnProGlnValSerProIleLysThrProSerVa 38
    5710 AATACGTTCTGACCCAACGAAACCCCTGCGCCAATCCGGGCACCTGCACA 5661

      38 lLeuIleThrLys.AspLysIleGlyAspHisHisThrHisGluHisAsp 54
    5660 GATGTGCACGAAGTGCCAAACTCGGGAGTCGCTACCGATCTGCGCACCGT 5611

      55 GluSerValSerHisValGly.....LeuGlnAlaHi 65
    5610 CGTCGACGATCGCGCTGGGGTGCTGATAATAACTCATGGCTTGAGACTCA 5561

      65 sPheGluThrTrpLeu.....GlnM 72
    5560 CACTCGATTGTGGTTGTTCAAACAGCCGCTGCAATTGACGGCTCGCGAAA 5511

      72 etHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGln 88
    5510 TATCCATAGCGGCGAGCCCCCACCTCGTTAAGGTGATGGCTATCCTGGTA 5461

      89 SerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAl 105
    5460 AATAAGCTC.....TCCGTCCTGATAAGGAG 5435

     105 aArgSerTrpGlnAlaCysGly.....HisGluProTyrGlnLeuP 119
    5434 CATCGGCGAAGAAGGCGCTGGAAGAAAAATCCATGAACCGTAC..... 5392

     119 roProGluHisLeuTrpGly.....GlnIleVal 128
    5391 .....GTTTGGAATACCCTCGGCTAGAGCACGCAACAGCTGGTT 5353

     129 ProThrLeuHisLeuTyrGlnAspLeuLysSer...ArgGlyIleLeuPr 144
    5352 CGCACCTTGCCAGCTGGAGCTAACAAGCGGAGCTGACAAACCCAGCTCCC 5303

     144 oAlaAsnThrGlnIleArgSerVal.....TyrArgAsnP 156
    5302 TGAAACGACGCACACGCTGCACGTTTGATTCGAACATCGGTATCTGCCCG 5253

     156 roGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
    5252 AGTAGAGCGACCTGCTTGCCGGCATAGCTGGTATCGACAAGGAAGGCACG 5203

     173 AsnSer.....AlaIleAspIleTrpValPro..... 181
    5202 CATAGCCTGGGCAAATGCCGGACTCTGCATCTGATACTGCCACATGCCCG 5153

     182 .....AspLeuGluIleLysSerGln..... 188
    5152 CCACAATGATCTT...GTCAAAGTTCAACATTGATTGGGCAACTGCATCA 5106

     189 .....AlaLeuTyrGluLeu 193
    5105 ATCTGCGCTTGGCAGGGTTTCCGCGCCCAACGGGGCAAACGTTCAAGATC 5056

     194 GlnAsnArgLeuCysGlnTyrTrp..... 201
```

```
    5055 GAAAGCAGG.........TATTGGCACACAACTGCTTCCGGTGAGTACTC 5015

     202 .LeuGluHis..................................... 204
         ::::::|||
    5014 GGTAAGCCACACCTGACTCGTTGCCAACCACGTCGAAGAAGTAGTTAAGC 4965

     205 ................GlyGluAsnGlnAsnPheGlyLeuGlyLeuTyr 215
                         :::::::|||:::   |||   |||:::
    4964 TGCGCAGCATGGCTATCACCGATCACCAGCACACGGGGTACGGCATTGAC 4915

     216 AlaThrGlyAlaIle......HisLeuAspThrGlnGlyPheArgLysTr 230
         |||::::::           :::|||:::|||::::::   |||   ||
    4914 GCTGCCTCGCTTGCACTCCCCTACCTGAACACCATGGCAGATCGACTCTG 4865

     230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
         |   :::         :::   |||         ||||||
    4864 GTACAGCGTAGCGCGTCAACTGAATTGGCGCCGGCGCCAC 4825

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-440-845D-8

seq_documentation_block:
; Sequence 8, Application US/08440845D
; Patent No. 5955329
;  GENERAL INFORMATION:
;    APPLICANT:  Yuan, L.
;    APPLICANT:  Kridl, J.
;    APPLICANT:  Dehesh, K.
;    APPLICANT:  Knauf, V.
;    TITLE OF INVENTION:  Engineering Plant Thioesterases For
;    TITLE OF INVENTION:  Altered Substrate Specificity.
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calgene, Inc.
;      STREET:  1920 Fifth Street
;      CITY:  Davis
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  95616
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.40 MB
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/440,845D
;      FILING DATE:  15-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donna E. Scherer
;      REGISTRATION NUMBER:  34,719
;      NAME:  Carl J. Schwedler
;      REGISTRATION NUMBER:  36,924
;      REFERENCE/DOCKET NUMBER:  CGNE 113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (916) 753-6313
;      TELEFAX:  (916) 753-1510
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1300 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-440-845D-8


alignment_scores:
            Quality:   79.50            Length:    244
              Ratio:  0.787              Gaps:     13
 Percent Similarity:  41.393   Percent Identity: 22.951

alignment_block:
 US-09-674-779-2 x US-08-440-845D-8/rev   ..
```

```
Align seq 1/1  to reverse of: US-08-440-845D-8  from: 1  to: 1300

     32 ProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHi 48
        |||::::::   |||      ::::::      ::::::   :::   ||
    645 CCTCTGGAAATGCCAATCTGAGTTCTCTTGGACAGTGAACCAAGTACTCA 596

     48 sHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaH 65
        ||||   |||:::|||
    595 TCACGAACATCAACAT.............................CGAC 576

     65 isPheGlu...ThrTrpLeu............................. 70
          ||||||   :::|||
    575 TTTTTGAAGTCGCCTGGTGTCTTGGTTCATCATTACCCACTTGCTAGTAG 526

     71 .................GlnMetHisHisAlaThrLysGlnGluValVa 81
                         |||   |||      ::::::::::::   ||
    525 CTCGGCCAATAACTTGACCAGTGGCATAGTCTCTCAGAATCCAATCACGT 476

     81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
        |   ::::::   :::|||:::             :::   :::||||||
    475 CTGGTTCCGATTTTTCCTTCACCCTGGCCCCACGACTCTATTTCCACCAC 426

     98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.GlyHisGl 114
           :::   |||      :::|||          ||||||:::   ::
    425 ATCACTCCAAGCTGGATATTTGTAGATTTCGATGTGCATGCGAGCAGTAA 376

    114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
        :|||||||||          :::|||      |||         ||||||
    375 CCCATATCAGACGCAATTTTCTCATGGTAGGGGTTGTCGAAAACCCACCC 326

    131 ..............................LeuHisLeuTyrGlnAsp 136
                                        :::|||   |||:::
    325 GTCGAATATCCAACGCTTTGGGCGTGATTGCATCCAACCTCCTGCAAGAG 276

    137 ............LeuLysSerArgGlyIleLeuProAlaAsnThrGlnIl 149
                    |||:::|||   |||::::::      |||      ||
    275 GTTGGCAATAGTCTCAACAGTAGCGGTCTTGTTAATCCCAACCTC...AT 229

    149 eArgSerValTyr..............ArgAsnProGluLeuAsnGlnC 161
        |   |||   |||               |||:::|||   |||:::|||:
    228 AGCATCTCACTATGAACTTCTCCTTATACGAAAGCCCGTCCTCGGTCAAG 179

    161 ysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp 177
        ::|||           :::|||:::         |||
    178 CTGCCCAGCCGGAGCCGGTCAGCCAGCCCCGAAGAAAC............ 141

    178 IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGl 194

    141 ............................................... 141

    194 nAsnArgLeuCysGlnTyrTrpLeuGluHis.........GlyGluAsnG 208
         :::|||||||||      |||::::::            |||   |||:
    140 .CACCGCCTCTGTTTTGAGTGGATTAACCTTGCTGGATGAGGAGGAAACC 92

    208 lnAsnPhe............Gly.LeuGlyLeuTyrAlaThrGlyAlaIl 220
        ::|||:::            ||| |||||||||          |||   ::
     91 ACAATTATCCGGGGAGGAATGGAGTTGGGCCTG.........GGCCGGGT 51

    220 eHisLeuAspThrGlnGlyPheArgLys 229
        :   :::   ::::::|||:::|||:::
     50 GGGAATGCGGGCCAATGGGCTTCGGGAA 23

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-868-458-8

seq_documentation_block:
; Sequence 8, Application US/08868458
; Patent No. 6150512
;   GENERAL INFORMATION:
;        APPLICANT: Yuan, L.
;        TITLE OF INVENTION: Engineering Plant Thioesterases And
;                            Disclosure of Plant Thioesterases
;                            Having No. 6150512el Substrate Specificity
```

```
;        NUMBER OF SEQUENCES: 19
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Calgene, Inc.
;             STREET: 1920 Fifth Street
;             CITY: Davis
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 95616
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
;             COMPUTER: IBM PC
;             OPERATING SYSTEM: Windows NT 4.0
;             SOFTWARE: Microsoft Word For Window 95 7.0a
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/868,458
;             FILING DATE: 03-Jun-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/US96/07064
;             FILING DATE: 15-MAY-96
;             APPLICATION NUMBER: 08/537,083
;             FILING DATE: 29-SEPT-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Carl J. Schwedler
;             REGISTRATION NUMBER: 36,924
;             REFERENCE/DOCKET NUMBER: CGNE 133
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (916) 753-6313
;             TELEFAX: (916) 753-1510
;    INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1300 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA to mRNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-868-458-8


alignment_scores:
            Quality:   79.50              Length:    244
              Ratio:  0.787               Gaps:       13
 Percent Similarity:  41.393   Percent Identity: 22.951

alignment_block:
 US-09-674-779-2 x US-08-868-458-8/rev   ..

 Align seg 1/1  to reverse of: US-08-868-458-8  from: 1  to: 1300

      32 ProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHi 48
         |||::::::   |||      ::::::      ::::::   :::   ||
     645 CCTCTGGAAATGCCAATCTGAGTTCTCTTGGACAGTGAACCAAGTACTCA 596

      48 sHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaH 65
         ||||   |||:::|||
     595 TCACGAACATCAACAT.............................CGAC 576

      65 isPheGlu...ThrTrpLeu............................. 70
         ||||||   :::|||
     575 TTTTTGAAGTCGCCTGGTGTCTTGGTTCATCATTACCCACTTGCTAGTAG 526

      71 .................GlnMetHisHisAlaThrLysGlnGluValVa 81
                          |||   |||      ::::::::::::   ||
     525 CTCGGCCAATAACTTGACCAGTGGCATAGTCTCTCAGAATCCAATCACGT 476

      81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
         |   ::::::   :::|||:::            :::   :::||||||
     475 CTGGTTCCGATTTTTCCTTCACCCTGGCCCCACGACTCTATTTCCACCAC 426

      98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.GlyHisGl 114
            :::   |||      :::|||         |||||| :::   ::
     425 ATCACTCCAAGCTGGATATTTGTAGATTTCGATGTGCATGCGAGCAGTAA 376

     114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
         :|||||||||         :::|||      |||          ||||||
     375 CCCATATCAGACGCAATTTTCTCATGGTAGGGGTTGTCGAAAACCCACCC 326

     131 ..............................LeuHisLeuTyrGlnAsp 136
                                        :::|||   |||:::
     325 GTCGAATATCCAACGCTTTGGGCGTGATTGCATCCAACCTCCTGCAAGAG 276

     137 ............LeuLysSerArgGlyIleLeuProAlaAsnThrGlnIl 149
                     |||:::|||   |||::::::      |||      ||
     275 GTTGGCAATAGTCTCAACAGTAGCGGTCTTGTTAATCCCAACCTC...AT 229

     149 eArgSerValTyr..............ArgAsnProGluLeuAsnGlnC 161
         |   |||   |||               |||:::|||   |||:::|||:
     228 AGCATCTCACTATGAACTTCTCCTTATACGAAAGCCCGTCCTCGGTCAAG 179

     161 ysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp 177
         ::|||            :::|||:::         |||
     178 CTGCCCAGCCGGAGCCGGTCAGCCAGCCCCGAAGAAAC............ 141

     178 IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGl 194

     141 ............................................... 141

     194 nAsnArgLeuCysGlnTyrTrpLeuGluHis.........GlyGluAsnG 208
          :::|||||||||      |||::::::            |||   |||:
     140 .CACCGCCTCTGTTTTGAGTGGATTAACCTTGCTGGATGAGGAGGAAACC 92

     208 lnAsnPhe............Gly.LeuGlyLeuTyrAlaThrGlyAlaIl 220
         ::|||:::            ||| |||||||||          |||   ::
      91 ACAATTATCCGGGGAGGAATGGAGTTGGGCCTG.........GGCCGGGT 51

     220 eHisLeuAspThrGlnGlyPheArgLys 229
         :   :::   ::::::|||:::|||:::
      50 GGGAATGCGGGCCAATGGGCTTCGGGAA 23

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-801-263A-8

seq_documentation_block:
; Sequence 8, Application US/08801263A
; Patent No. 5811407
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 5811407th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,263A
;      FILING DATE:  19-FEB-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  8:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11703 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-801-263A-8


alignment_scores:
            Quality:   79.50              Length:    211
              Ratio:  0.837                Gaps:     14
 Percent Similarity:  45.024   Percent Identity: 25.118

alignment_block:
 US-09-674-779-2 x US-08-801-263A-8   ..

 Align seg 1/1  to: US-08-801-263A-8  from: 1  to: 11703

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||       |||   :::||||||             |||::
    8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
         :|||      ||||||:::|||      :::|||             |||:
    8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 8739

      44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
         ::      |||  ||||||||||||         :::   |||
    8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 8782

      60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                               |||:::::::      :::::::::
    8783 .........................TGGATACGACCAAAGCGGAGCAGC 8806

      77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
         |||::::::   ||| ||||||::::::    |||
    8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTGAGCAGGATCACACCGTTA 8856

      88 ................................................... 88

    8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

      89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
               |||      |||   :::|||             ||||||
    8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

      98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
             ::::::     ::::::::::::::::|||      :::|||
    8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

     112 ............................GlyHisGluProTyrGlnLeu 118
                                     |||:::|||   |||:::|||
    9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
         ||||||   |||      |||:::   :::|||         :::|||::
    9057 CCTCCCGTTCAC......GGTAAAAAAATTCCTTGCACA...GTGTACGA 9097

     135 nAspLeuLys......SerArgGlyIleLeu............ProAla. 145
         :   ||||||      :::   |||   :::             |||:::
    9098 CCGTCTGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGAC 9147

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
         :::|||   |||:::   ::::::::::|||   :::
    9148 CGCACGCTTATACATCCTACCTGGAAGAATCATCAG.............. 9183

     162 aGlyGlyAlaAlaMetSerLysHisLeu 171
          |||      :::::::|||:::|||||||
    9184 .GGAAAGTTTACGCAAAGCCGCCATCTG 9210

 seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-102-248-8

seq_documentation_block:
; Sequence 8, Application US/09102248
; Patent No. 6008035
;  GENERAL INFORMATION:
;    APPLICANT:  Johnston, Robert E.
;    APPLICANT:  Davis, Nancy L.
;    APPLICANT:  Simpson, Dennis A.
;    TITLE OF INVENTION:  System for the In Vivo Delivery and
;    TITLE OF INVENTION:  Expression of Heterologous Genes in the Bone Marrow
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bell Seltzer Park & Gibson, P.A.
;      STREET:  1211 East Morehead Street
;      CITY:  Charlotte
;      STATE:  No. 6008035th Carolina
;      COUNTRY:  USA
;      ZIP:  28234
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/102,248
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/801,263
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sibley, Kenneth D.
;      REGISTRATION NUMBER:  31,665
;      REFERENCE/DOCKET NUMBER:  5470-147
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-420-2200
;      TELEFAX:  919-881-3175
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11703 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-102-248-8


alignment_scores:
            Quality:   79.50              Length:    211
              Ratio:  0.837                Gaps:     14
 Percent Similarity:  45.024   Percent Identity: 25.118

alignment_block:
 US-09-674-779-2 x US-09-102-248-8   ..

 Align seg 1/1  to: US-09-102-248-8  from: 1  to: 11703

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||       |||   :::||||||             |||::
    8649 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8698

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
         :|||      ||||||:::|||      :::|||             |||:
    8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 8739

      44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
         ::      |||  ||||||||||||         :::   |||
    8740 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 8782

      60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                               |||:::::::      :::::::::
    8783 .........................TGGATACGACCAAAGCGGAGCAGC 8806

      77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln.............. 88
```

```
      |||::::::   ||| ||||||::::::      |||
 8807 AAGCGCAAACAAGTACCGCTACATGTCGCTTGAGCAGGATCACACCGTTA 8856

   88 ..................................................  88

 8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

   89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
         |||      |||   :::|||             ||||||
 8907 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 8956

   98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
        ::::::      :::::::::::::::|||      :::|||
 8957 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9006

  112 ............................GlyHisGluProTyrGlnLeu 118
                                 |||:::|||   |||:::|||
 9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056

  119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
      ||||||   |||      |||:::   :::|||         :::|||::
 9057 CCTCCCGTTCAC......GGTAAAAAAATTCCTTGCACA...GTGTACGA 9097

  135 nAspLeuLys......SerArgGlyIleLeu............ProAla. 145
      :   ||||||     :::   |||   :::             |||:::
 9098 CCGTCTGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGAC 9147

  146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
      :::|||   |||:::   ::::::::::|||   :::
 9148 CGCACGCTTATACATCCTACCTGGAAGAATCATCAG.............. 9183

  162 aGlyGlyAlaAlaMetSerLysHisLeu 171
       |||      ::::::|||:::|||||
 9184 .GGAAAGTTTACGCAAAGCCGCCATCTG 9210
```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-00362-4

```
seq_documentation_block:
; Sequence 4, Application PC/TUS9500362
;  GENERAL INFORMATION:
;     APPLICANT:  IMMUNEX CORPORATION
;     TITLE OF INVENTION:  Ligand That Binds Fas Antigen
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Immunex Corporation
;       STREET:  51 University Street
;       CITY:  Seattle
;       STATE:  WA
;       COUNTRY:  US
;       ZIP:  98101
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  Apple 7.1
;       SOFTWARE:  Microsoft Word, Version 5.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/00362
;       FILING DATE:  06-JAN-1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/179,138
;       FILING DATE:  07-JAN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/190,559
;       FILING DATE:  01-FEB-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Anderson, Kathryn A.
;       REGISTRATION NUMBER:  32,172
;       REFERENCE/DOCKET NUMBER:  2805-WO
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1171 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  31..870
PCT-US95-00362-4


alignment_scores:
            Quality:   79.00              Length:    250
              Ratio:  0.814                Gaps:     15
 Percent Similarity:  38.800   Percent Identity: 25.200

alignment_block:
 US-09-674-779-2 x PCT-US95-00362-4    ..

 Align seg 1/1  to: PCT-US95-00362-4  from: 1  to: 1171

   23 ProIleProThrAsnProGlnValSerProIleLysThrProSer..... 37
      |||   |||      |||   ||||||||||:::      ||||||
  163 CCGCCACCTCCACCACCACCTGTGTCACCACTACCACCGCCATCACAACC 212

   38 ..........ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
                ::|||      |||      ||||||:::|||:
  213 ACTCCCACTGTCGCCACTGACCCCTCTAAAGAAGAAGGACCACAACACAA 262

   51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
      ::
  263 AT.................................................. 264

   68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
         ||||||                          |||   :::
  265 CTGTGGCTACCG.......................GTGGTATTTTT 287

   84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnL 101
            |||   :::   :::|||      |||      |||   ||||
  288 CATGGTTCTGGTGGGTCTGGTTGGAATGGGATTA...GGAATGTATCAGC 334

  101 euLeu.ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
      ||::: |||      |||:::||||||            :::|||   ::
  335 TCTTCCACCTGCAGAAGGAACTGGCAGAACTCCGTGAGTTCACCAACCAA 384

  117 nLeuProProGluHisLeuTrpGlyGlnIleValProThr...LeuHisL 133
      :           ||||||      :::      ||||||   :::|||
  385 AGCCTTAAAGTATCATCTTTTGAAAAGCAAATAGCCAACCCCAGTACACC 434

  133 euTyrGlnAspLeuLysSerArgGlyIleLeuProAla.AsnThrGlnIl 149
         :::      ||||||||||||:::   |||    |||
  435 CTCTGAAAAAA...AAGAGCCGAGGAGTGTGGCCCATTTAAC........ 473

  149 eArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaA 166
                 |||:::|||   |||:::         :::|||
  474 .............AGGGAACCCCCACTCAAGGTCCATCCCTCTGGA.... 506

  166 laMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp......... 179
        |||::::::|||:::   |||   :::   |||:::|||
  507 ..ATGGGAAGACACATATGGAACCGCTCT...GATCTCTGGAGTGAAGTA 551

  179 ..................................................  179

  552 TAAGAAAGGTGGCCTTGTGATCAACGAAACTGGGTTGTACTTCGTGTATT 601

  180 ..........ValProAspLeuGluIle..............LysSerG 188
                ::|||         :::                |||   |
  602 CCAAAGTATACTTCCGGGGTCAGTCTTGCAACAACCAGCCCCTAAACCAC 651

  188 lnAlaLeuTyrGlu..................................... 192
      ||:::||||||||||
  652 AAGGTCTATATGAGGAACTCTAAGTATCCTGAGGATCTGGTGCTAATGGA 701
```

```
     193 ...LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGl 208
     702 GGAGAAGAGGTTGAACTACTTCACTACTGGACAGATATGGGCCCACAGCA 751

     208 n.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHisLeu 222
     752 GCTACCTGGGGGCAGTATTCAATCTTACCAGTGCTGACCATTTA 795
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-165-241-2

seq_documentation_block:
; Sequence 2, Application US/09165241
; Patent No. 6130077
; GENERAL INFORMATION:
; APPLICANT: Henry Yue
; APPLICANT: Karl J. Guegler
; APPLICANT: Neil C. Corley
; APPLICANT: Janice Au-Young
; TITLE OF INVENTION: HUMAN CYTOCHROME P450
; FILE REFERENCE: PF-0608 US
; CURRENT APPLICATION NUMBER: US/09/165,241
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 991729
US-09-165-241-2

alignment_scores:
Quality: 78.00 Length: 213
Ratio: 0.750 Gaps: 10
Percent Similarity: 48.826 Percent Identity: 23.005

alignment_block:
US-09-674-779-2 x US-09-165-241-2 ..

Align seg 1/1 to: US-09-165-241-2 from: 1 to: 1648

```
      67 GluThrTrpLeu......GlnMetHisHisAlaThrLysGlnGluValVa 81
       9 GAGACGTGGCTCCCTGGGCGGCAGAACCATGTTGGACTTCGCGATCTTCG 58

      81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
      59 CCGTTAC......CTTCTTGCTGGCGTTGGTGGGAGCCGTGCTCTACCTC 102

      98 etSerGlnLeuLeuThrThrAlaArgSer..................... 107
     103 TATCCGGCTTCCAGACAAGCTGCAGGAATTCCAGGGATTACTCCAACTGA 152

     108 .................................TrpGln........ 109
     153 AGAAAAAGATGGTAATCTTCCAGATATTGTGAATAGTGGAAGTTTGCATG 202

     110 ............................AlaCysGlyHisGluProT 116
     203 AGTTCCTGGTTAATTTGCATGAGAGATATGGGCCTGTGGTCTCCTTCTGG 252

     116 yrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHis 132
     253 TTTGGCAGGCGCCTCGTGGTTAGTTTGGGCACTGTTGATGTACTGAAGCA 302

     133 LeuTyrGlnAspLeu.LysSerArgGlyIleLeuProAlaAsnThrGlnI 149
     303 GCATATCAATCCCAATAAGACA......TTGGACCCTTTTGAAACCATGC 346

     149 leArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla 165
```

```
     347 TGAAGTCATTATTAAGG............TATCAATCTGGTGGTGGCAGT 384

     166 AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAs 182
     385 GTGAGTGAAAACCACATGAGGAAAAAATTGTATGAAAATGGTGTGACTGA 434

     182 p...LeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuC 198
     435 TTCTCTGAAGAGTAACTTTGCCCTCCTCCTAAAGCTTTCAGAAGAATTAT 484

     198 ysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu 214
     485 TAGATAAATGGCTCTCCTACCCAGAGACCCAGCACGTGCCCCTCAGCCAG 534

     215 TyrAlaThrGly...AlaIleHisLeuAspThrGlnGlyPheArgLysTr 230
     535 CATATGCTTGGTTTTGCTATGAAGTCTGTTACACAG......ATGGTAAT 578

     230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArg 242
     579 GGGTAGTACATTTGAAGATGATCAGGAAGTCATTCGC 615
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-120-1

seq_documentation_block:
; Sequence 1, Application US/08458120
; Patent No. 5635616
; GENERAL INFORMATION:
; APPLICANT: OLSEN, ET AL.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,120
; FILING DATE: SUBMITTED HEREWITH
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05561
; FILING DATE: 18 MAY 94
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: cDNA
US-08-458-120-1

alignment_scores:
Quality: 78.00 Length: 202
Ratio: 0.857 Gaps: 11
Percent Similarity: 45.050 Percent Identity: 24.257

```
alignment_block:
 US-09-674-779-2 x US-08-458-120-1   ..

 Align seg 1/1  to: US-08-458-120-1  from: 1  to: 1882

      82 ArgTyrGln.........AlaTyrLeuGln..............SerAr 90
         |||||||||          ::::::|||                :::::
     889 AGATATCAGCCAGATGTGAAGTTTATCCAGGTTGATATCTGTGCAGAAGA 938

      90 gLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu.............. 102
         :|||||||||   :::   |||         ||||||
     939 ATTGGGGAATAATGTAAAGCCCGCTGTTACTTTGCTAGGAAACATACATG 988

     103 ..ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu 118
           :::|||   :::      :::               |||:::|||
     989 CTGTCACTAAGCAGCTTTTAGAGGAACTTGATAAAACACCATGGCAGTAT 1038

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHis........ 132
         |||||||||      |||   :::         |||||||::
    1039 CCTCCAGAGAGCAAGTGGTGGAAA.........ACTCTGAGAGAAAAAAT 1079

     133 ................LeuTyrGlnAspLeuLysSerArgGlyIleLeuP 144
                         :::::::|||   |||:::      ||||
    1080 GAAGAGCAATGAAGCTGCATCCAAGGAACTAGCTTCTAAAAAATCCCTGC 1129

     144 roAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu...Asn 159
         ||   |||                  :::   :::   :::|||   :::
    1130 CTATGAATTATTACACAGTATTCTACCATGTTCAAGAACAACTACCTAGA 1179

     160 GlnCys.........AlaGlyGlyAlaAlaMetSerLysHisLeuThrAs 173
         :::|||         :::   ||||||
    1180 GACTGTTTCGTGGTAAGTGAAGGAGCA...................... 1206

     173 nSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaL 190
          ::::::::::||||||                            :::
    1207 .AATACTATGGACATT............................GGAC 1225

     190 euTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGlu 206
               |||||||||   |||   ::::::   |||:::   |||
    1226 GGACTGTGCTTCAGAACTACCTTCCTCGTCACAGGCTTGATGCTGGTACT 1275

     207 AsnGlnAsnPheGlyLeuGlyLeu.........TyrAlaThrGlyAlaIl 220
           ::::::|||:::||||||           |||:::::::::::
    1276 TTCGGAACAATGGGAGTTGGTTTGGGATTTGCTATTGCAGCTGCCGTGGT 1325

     220 eHisLeuAsp.........................ThrGlnGlyPheA 228
         :     |||                         :::::|||
    1326 GGCTAAAGATAGAAGCCCTGGGCAATGGATCATCTGTGTGGAAGGAGACA 1375

     228 rgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisVal 244
           :::|||   :::      |||::::::::::|||||||||:::
    1376 GTGCATTTGGGTTTTCTGGCATGGAGGTAGAAACCATCTGCAGGTACAAC 1425

     245 LeuPro 246
         ||||||
    1426 TTGCCA 1431

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-867-970-1

seq_documentation_block:
; Sequence 1, Application US/08867970
; Patent No. 5945273
;  GENERAL INFORMATION:
;    APPLICANT:  Olsen, Henrik S.
;    APPLICANT:  Coleman, Timothy
;    APPLICANT:  Adams, Mark D.
;    TITLE OF INVENTION:  Human Oxalyl-CoA Decarboxylase
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,970
;      FILING DATE:  JUN-3-97
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/458,120
;      FILING DATE:  02-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US94/05561
;      FILING DATE:  18-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Robert H.
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PF119D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-309-8504
;      TELEFAX:  301-309-8439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1882 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..1744
US-08-867-970-1


alignment_scores:
           Quality:   78.00              Length:    202
             Ratio:  0.857                Gaps:     11
 Percent Similarity:  45.050   Percent Identity: 24.257

alignment_block:
 US-09-674-779-2 x US-08-867-970-1   ..

 Align seg 1/1  to: US-08-867-970-1  from: 1  to: 1882

      82 ArgTyrGln.........AlaTyrLeuGln..............SerAr 90
         |||||||||          ::::::|||                :::::
     889 AGATATCAGCCAGATGTGAAGTTTATCCAGGTTGATATCTGTGCAGAAGA 938

      90 gLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu.............. 102
         :|||||||||   :::   |||         ||||||
     939 ATTGGGGAATAATGTAAAGCCCGCTGTTACTTTGCTAGGAAACATACATG 988

     103 ..ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu 118
           :::|||   :::      :::               |||:::|||
     989 CTGTCACTAAGCAGCTTTTAGAGGAACTTGATAAAACACCATGGCAGTAT 1038

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHis........ 132
         |||||||||      |||   :::         |||||||::
    1039 CCTCCAGAGAGCAAGTGGTGGAAA.........ACTCTGAGAGAAAAAAT 1079

     133 ................LeuTyrGlnAspLeuLysSerArgGlyIleLeuP 144
                         :::::::|||   |||:::      ||||
    1080 GAAGAGCAATGAAGCTGCATCCAAGGAACTAGCTTCTAAAAAATCCCTGC 1129

     144 roAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu...Asn 159
         ||   |||                  :::   :::   :::|||   :::
    1130 CTATGAATTATTACACAGTATTCTACCATGTTCAAGAACAACTACCTAGA 1179

     160 GlnCys.........AlaGlyGlyAlaAlaMetSerLysHisLeuThrAs 173
         :::|||         :::   ||||||
```

```
    1180 GACTGTTTCGTGGTAAGTGAAGGAGCA..................... 1206

     173 nSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaL 190
          :::::::::||||||                              :::
    1207 .AATACTATGGACATT..........................GGAC 1225

     190 euTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGlu 206
              |||||||||   |||   ::::::   |||:::   |||
    1226 GGACTGTGCTTCAGAACTACCTTCCTCGTCACAGGCTTGATGCTGGTACT 1275

     207 AsnGlnAsnPheGlyLeuGlyLeu.........TyrAlaThrGlyAlaIl 220
              ::::::|||:::||||||           |||::::::::::::
    1276 TTCGGAACAATGGGAGTTGGTTTGGGATTTGCTATTGCAGCTGCCGTGGT 1325

     220 eHisLeuAsp.........................ThrGlnGlyPheA 228
          :     |||                         :::::|||
    1326 GGCTAAAGATAGAAGCCCTGGGCAATGGATCATCTGTGTGGAAGGAGACA 1375

     228 rgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisVal 244
              :::|||   :::      |||::::::::::||||||||:::
    1376 GTGCATTTGGGTTTTCTGGCATGGAGGTAGAAACCATCTGCAGGTACAAC 1425

     245 LeuPro 246
         ||||||
    1426 TTGCCA 1431

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-326-217-1

seq_documentation_block:
; Sequence 1, Application US/09326217
; Patent No. 6200796
;  GENERAL INFORMATION:
;    APPLICANT:  Olsen, Henrik S.
;    APPLICANT:  Coleman, Timothy
;    APPLICANT:  Adams, Mark D.
;    TITLE OF INVENTION:  Human Oxalyl-CoA Decarboxylase
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/326,217
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,970
;      FILING DATE:  JUN-03-97
;      APPLICATION NUMBER:  US 08/458,120
;      FILING DATE:  02-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US94/05561
;      FILING DATE:  18-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Robert H.
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PF119D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-309-8504
;      TELEFAX:  301-309-8439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1882 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..1744
US-09-326-217-1


alignment_scores:
            Quality:   78.00              Length:    202
              Ratio:  0.857                Gaps:     11
 Percent Similarity:  45.050   Percent Identity: 24.257

alignment_block:
 US-09-674-779-2 x US-09-326-217-1   ..

 Align seg 1/1  to: US-09-326-217-1  from: 1  to: 1882

      82 ArgTyrGln.........AlaTyrLeuGln..............SerAr 90
         |||||||||          :::::|||                 :::::
     889 AGATATCAGCCAGATGTGAAGTTTATCCAGGTTGATATCTGTGCAGAAGA 938

      90 gLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu..... 102
         :|||||||||   :::   |||         ||||||
     939 ATTGGGGAATAATGTAAAGCCCGCTGTTACTTTGCTAGGAAACATACATG 988

     103 ..ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu 118
           :::|||   :::       :::               |||:::|||
     989 CTGTCACTAAGCAGCTTTTAGAGGAACTTGATAAAACACCATGGCAGTAT 1038

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHis........ 132
         |||||||||      |||   :::         ||||||:::
    1039 CCTCCAGAGAGCAAGTGGTGGAAA.........ACTCTGAGAGAAAAAAT 1079

     133 ..............LeuTyrGlnAspLeuLysSerArgGlyIleLeuP 144
                        :::::::|||   |||:::      ||||
    1080 GAAGAGCAATGAAGCTGCATCCAAGGAACTAGCTTCTAAAAAATCCCTGC 1129

     144 roAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu...Asn 159
         ||   |||                  :::   :::   :::|||   :::
    1130 CTATGAATTATTACACAGTATTCTACCATGTTCAAGAACAACTACCTAGA 1179

     160 GlnCys.........AlaGlyGlyAlaAlaMetSerLysHisLeuThrAs 173
         :::|||         :::   ||||||
    1180 GACTGTTTCGTGGTAAGTGAAGGAGCA..................... 1206

     173 nSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaL 190
          :::::::::||||||                              :::
    1207 .AATACTATGGACATT..........................GGAC 1225

     190 euTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGlu 206
              |||||||||   |||   ::::::   |||:::   |||
    1226 GGACTGTGCTTCAGAACTACCTTCCTCGTCACAGGCTTGATGCTGGTACT 1275

     207 AsnGlnAsnPheGlyLeuGlyLeu.........TyrAlaThrGlyAlaIl 220
              ::::::|||:::||||||           |||::::::::::::
    1276 TTCGGAACAATGGGAGTTGGTTTGGGATTTGCTATTGCAGCTGCCGTGGT 1325

     220 eHisLeuAsp.........................ThrGlnGlyPheA 228
          :     |||                         :::::|||
    1326 GGCTAAAGATAGAAGCCCTGGGCAATGGATCATCTGTGTGGAAGGAGACA 1375

     228 rgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisVal 244
              :::|||   :::      |||:::::::::|||||||||:::
    1376 GTGCATTTGGGTTTTCTGGCATGGAGGTAGAAACCATCTGCAGGTACAAC 1425

     245 LeuPro 246
         ||||||
    1426 TTGCCA 1431

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-972-218-1

seq_documentation_block:
```

```
; Sequence 1, Application US/08972218
; Patent No. 6197502
;  GENERAL INFORMATION:
;    APPLICANT:  Renner, Wolfgang A.
;    APPLICANT:  Orberger, Georg H.
;    APPLICANT:  Koller, Daniel
;    APPLICANT:  Bailey, James E.
;    TITLE OF INVENTION:  EXPRESSION CLONING PROCESSES FOR THE
;    TITLE OF INVENTION:  DISCOVERY, CHARACTERIZATION AND ISOLATION OF GENES
;    TITLE OF INVENTION:  ENCODING POLYPEPTIDES WITH A PREDETERMINED PROPERTY
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of The Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2811
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/972,218
;      FILING DATE:  14-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Abrams, Samuel
;      REGISTRATION NUMBER:  30,605
;      REFERENCE/DOCKET NUMBER:  008358-0004-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-493-4935
;      TELEFAX:  650-493-5556
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13905 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-972-218-1


alignment_scores:
            Quality:   77.50            Length:    211
              Ratio:  0.816              Gaps:     14
 Percent Similarity:  45.024   Percent Identity: 25.118

alignment_block:
 US-09-674-779-2 x US-08-972-218-1   ..

 Align seg 1/1  to: US-08-972-218-1  from: 1  to: 13905

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         ||||||   |||      |||   :::||||||           |||::
    8916 ACCCTGACCAGCCCCTACTTGGGCACATGCTCGTACTGCCACCATACTGA 8965

      27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
         :|||      ||||||:::|||      :::|||            |||:
    8966 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.........GACG 9006

      44 ysIleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
         ::      |||  ||||||||||||         :::   |||
    9007 AAGCGGACGATAACACCATACGCATACAGACTTCCGCCCAGTT....... 9049

      60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThr 76
                                |||::::::      :::::::::
    9050 .........................TGGATACGACCATAGCGGAGCAGC 9073

      77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln..... 88
         |||::::::   ||| ||||||::::::      |||
    9074 AAGCGCAAACAAGTACCGCTACATGTCGCTTAAGCAGGATCACACCGTTA 9123

      88 ..... 88

    9124 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 9173

      89 ......SerArgLeuGlyAsnTyrLeu............ProPro..... 97
               |||      |||   :::|||            ||||||
    9174 AGGCTTAGCTACAAAGGATACTTTCTCCTCGCAAAATGCCCTCCAGGGGA 9223

      98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
           ::::::      ::::::::::::::::|||      :::|||
    9224 CAGCGTAACGGTTAGCATAGTGAGTAGCAACTCAGCAACGTCATGTACAC 9273

     112 ............................GlyHisGluProTyrGlnLeu 118
                                     |||:::|||   |||:::|||
    9274 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9323

     119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
         ||||||   |||      |||:::   :::|||         :::|||::
    9324 CCTCCCGTTCAC......GGTAAAAAAATTCCTTGCACA...GTGTACGA 9364

     135 nAspLeuLys......SerArgGlyIleLeu............ProAla. 145
         :   ||||||      :::   |||   :::            |||:::
    9365 CCGTCTGAAAGAAACAACTGCAGGCTACATCACTATGCACAGGCCGGGAC 9414

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
         :::|||   |||:::   :::::::::|||   :::
    9415 CGCACGCTTATACATCCTACCTGGAAGAATCATCAG.............. 9450

     162 aGlyGlyAlaAlaMetSerLysHisLeu 171
          |||      ::::::|||:::||||||
    9451 .GGAAAGTTTACGCAAAGCCGCCATCTG 9477
```

OM of: US-09-674-779-2 to: EST:* out_format : pfs

Date: Sep 16, 2002 6:20 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124121_25821/app_query.fasta_1.310
-DB=EST -QFMT=fastap -SUFFIX=std.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09674779_@CGN1_1_2604
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1624.540000

score_list:

| Sequence | Strd | Orig | ZScore | EScore | Len | ! Documentation .. |
|---|---|---|---|---|---|---|
| gb_gss:AG078683 | - | 110.50 | 185.32 | 0.4476 | 645 | ! AG078683 Pan troglodytes DNA, c |
| gb_gss:AZ638282 | + | 105.50 | 176.42 | 1.40 | 640 | ! AZ638282 1M0498P02F Mouse 10kb |
| gb_gss:AZ953717 | - | 103.50 | 172.97 | 2.18 | 632 | ! AZ953717 2M0219P02F Mouse 10kb |
| gb_gss:AZ175435 | - | 98.50 | 161.16 | 9.92 | 817 | ! AZ175435 SP_0132_A1_H03_T7A Str |
| gb_est2:BM467307 | + | 98.50 | 156.28 | 18.54 | 1270 | ! BM467307 AGENCOURT_6471965 NIH |
| gb_gss:AG140443 | + | 98.00 | 162.90 | 7.93 | 643 | ! AG140443 Pan troglodytes DNA, c |
| gb_est2:BG214490 | + | 97.00 | 161.49 | 9.51 | 621 | ! BG214490 RST34131 Athersys RAGE |
| gb_est2:BE273034 | - | 97.00 | 159.20 | 12.75 | 764 | ! BE273034 601171358F1 NIH_MGC_14 |
| gb_est2:BI454968 | + | 97.00 | 157.72 | 15.43 | 874 | ! BI454968 603173541F1 NCI_CGAP_M |
| gb_gss:CNS04OFM | - | 96.00 | 156.72 | 17.53 | 813 | ! AL299947 Tetraodon nigroviridis |
| gb_est2:BE442326 | + | 95.50 | 162.26 | 8.62 | 454 | ! BE442326 925017E08.x1 C. reinha |
| gb_gss:BH032955 | + | 95.00 | 155.47 | 20.59 | 774 | ! BH032955 RPCI-24-294G15.TV RPCI |
| gb_est2:BF627823 | - | 95.00 | 154.99 | 21.88 | 808 | ! BF627823 HVSMEb0005O09f Hordeum |
| gb_est2:BE974669 | - | 95.00 | 152.16 | 31.46 | 1044 | ! BE974669 601680707R2 NIH_MGC_8 |
| gb_est2:BI728822 | - | 94.50 | 157.40 | 16.07 | 599 | ! BI728822 1031101G02.y1 C. reinh |
| gb_est2:BE402495 | + | 94.50 | 157.29 | 16.29 | 605 | ! BE402495 CSB008E09F990908 ITEC |
| gb_est2:BG845267 | + | 93.50 | 153.88 | 25.22 | 700 | ! BG845267 1024009E05.x1 C. reinh |
| gb_est2:BG840148 | - | 93.00 | 155.65 | 20.11 | 550 | ! BG840148 MEST8-E10.T3 ISUM3-TL |
| gb_est2:BF578004 | + | 92.50 | 140.83 | 134.49 | 1939 | ! BF578004 602091838F1 NCI_CGAP_ |
| gb_est2:BE426799 | + | 92.00 | 152.34 | 30.76 | 631 | ! BE426799 WHE0332_C08_F16ZS Whea |
| gb_est1:BB821040 | + | 91.50 | 155.42 | 20.71 | 440 | ! BB821040 BB821040 RIKEN full-le |
| gb_est2:BJ124438 | + | 91.50 | 151.45 | 34.44 | 630 | ! BJ124438 BJ124438 unpublished c |
| gb_est1:AL536956 | - | 91.50 | 147.02 | 60.80 | 941 | ! AL536956 AL536956 LTI_FL013_FBr |
| gb_est2:BG848755 | + | 91.00 | 149.83 | 42.43 | 673 | ! BG848755 1024022F10.x2 C. reinh |
| gb_gss:CNS01Y2Y | - | 91.00 | 146.79 | 62.64 | 886 | ! AL172483 Tetraodon nigroviridis |
| gb_est2:BE779469 | - | 91.00 | 146.14 | 68.12 | 940 | ! BE779469 601464207F1 NIH_MGC_67 |
| gb_est2:BJ111789 | + | 90.50 | 149.30 | 45.42 | 651 | ! BJ111789 BJ111789 unpublished c |
| gb_est2:W80001 | - | 90.50 | 149.21 | 45.92 | 656 | ! W80001 me88h09.r1 Soares mouse |
| gb_est2:BG165572 | - | 90.50 | 147.47 | 57.41 | 768 | ! BG165572 602343723F1 NIH_MGC_89 |
| gb_gss:BH269012 | - | 90.50 | 146.92 | 61.58 | 807 | ! BH269012 CH230-66O5.TJ CHORI-23 |
| gb_gss:AQ744833 | - | 90.50 | 146.82 | 62.45 | 815 | ! AQ744833 HS_5506_A2_G06_SP6 RPC |
| gb_est2:BG292887 | + | 90.50 | 143.96 | 90.01 | 1055 | ! BG292887 602389522F1 NIH_MGC_9 |
| gb_gss:AZ083675 | - | 90.00 | 149.41 | 44.76 | 594 | ! AZ083675 RPCI-23-22M17.TJ RPCI- |
| gb_est1:BB665280 | - | 90.00 | 147.60 | 56.49 | 700 | ! BB665280 BB665280 RIKEN full-le |
| gb_est2:BG368308 | + | 90.00 | 146.89 | 61.82 | 746 | ! BG368308 HVSMEi0017J13f Hordeum |
| gb_gss:BH070893 | + | 89.00 | 148.81 | 48.34 | 533 | ! BH070893 RPCI-24-275C20.TV RPCI |
| gb_est1:BB692619 | - | 89.00 | 148.46 | 50.54 | 550 | ! BB692619 BB692619 RIKEN full-le |
| gb_est2:BG207133 | - | 89.00 | 146.67 | 63.61 | 647 | ! BG207133 RST26599 Athersys RAGE |
| gb_gss:CNS02K1D | + | 89.00 | 141.61 | 121.73 | 1023 | ! AL200938 Tetraodon nigroviridi |
| gb_est2:BG291582 | + | 89.00 | 141.51 | 123.25 | 1032 | ! BG291582 602385677F1 NIH_MGC_9 |
| gb_est2:BI993724 | + | 88.50 | 148.96 | 47.45 | 485 | ! BI993724 1031003D12.x2 C. re |
| gb_est2:BF177880 | + | 88.50 | 148.42 | 50.81 | 509 | ! BF177880 Ljirnpest36-887-d7 |
| gb_est2:BM002719 | + | 88.50 | 148.12 | 52.80 | 523 | ! BM002719 1031106C05.x1 C. re |
| gb_est1:AW757658 | + | 88.50 | 147.65 | 56.12 | 546 | ! AW757658 874002G08.x1 C. rei |
| gb_gss:AQ345896 | - | 88.50 | 147.29 | 58.76 | 564 | ! AQ345896 RPCI11-126I23.TJ RP |

seq_name: gb_gss:AG078683

seq_documentation_block:
LOCUS AG078683 645 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-074B02.F, genomic survey sequence.
ACCESSION AG078683
VERSION AG078683.1 GI:16630485
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-074B02.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 645)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .645
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-074B02.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 104 a 152 c 213 g 176 t
ORIGIN

alignment_scores:
Quality: 110.50 Length: 136
Ratio: 1.811 Gaps: 5
Percent Similarity: 44.853 Percent Identity: 26.471

alignment_block:
US-09-674-779-2 x AG078683/rev ..

Align seg 1/1 to reverse of: AG078683 from: 1 to: 645

```
      24 IleProThrAsnProGlnValSerProIle..................... 33
         :::||||||||||||||||:::   |||:::
     561 TTGCCAACAAACCCCCAAATCCCTCCATTGGCCTGTGGCAACCGTTGACG 512

      34 ..............LysThrProSerValLeu................... 39
                        :::||||||::::::::
     511 CAACTTTTGGCTGTAGCAAACCCCTCAGATAGTGGAAAAGAAACACACAG 462

      40 ....................................IleThrLysAsp 43
```

```
                                                  :::||||||||||
    461 AGACTGAGAGGCAGAGAGAAAAGAGAGAACGGGAGCCACTCACAAAAGAC 412

     44 LysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHisVa 60
                      |||||||||   |||   ::::::   :::|||
    411 ACCTACATACACACACACACACACACACACACACACACACACACACACTC 362

     60 lGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrL 77
       |||      ::::::   |||                :::      :::
    361 AGGGTCATACAGCAGAGCAGAG..............CAGCCCAGGGTCA 327

     77 ysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsn 93
        |||:::|||         ::::::      ::::::||||||   :::
    326 CGCAAGACGTCCCGTCATCGAGATCACGATCGATGGCACGACTTTGGGGG 277

     94 TyrLeuProProMetSerGlnLeuLeuThrThr.AlaArgSerTrpGlnA 110
         |||   |||                  ||| :::   |||   ||||
    276 ACACTCACTCCA.................ACCAACACCATCCGTGCAGG 245

    110 laCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGln 126
         ||   |||      ||||||   |||||||||:::      ||||||:::
    244 CCTCAGGCTGGTATCCCTATCTGCTACCCCCGTCTGCCGCCTGGGGTTCC 195

    127 IleVal 128
           :::
    194 ACCATC 189
```

seq_name: gb_gss:AZ638282

seq_documentation_block:

```
LOCUS       AZ638282               640 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0498P02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0498P02 F, DNA sequence.
ACCESSION   AZ638282
VERSION     AZ638282.1  GI:11760472
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0498  row: P  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 640.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0498P02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      181 a    181 c    122 g    155 t      1 others
ORIGIN
```

alignment_scores:

```
            Quality:  105.50              Length:    150
              Ratio:  1.353                 Gaps:      7
 Percent Similarity:  52.000   Percent Identity: 28.667
```

alignment_block:

US-09-674-779-2 x AZ638282 ..

Align seg 1/1 to: AZ638282 from: 1 to: 640

```
     35 ThrProSerValLeuIleThrLysAspLysIleGlyAspHis..... 48
        ||||||:::|||::::::                  |||
     34 ACACCTAAGGTTGTACTCTGACCTGTACACACACACACACACACACAC 83

     49 .HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaH 65
         |||||||||   |||   ::::::   :::|||   ::::::::::::|
     84 ACACACACACACACATACACACGCACGCACACTGTGCTGTACACACAC 133

     65 isPheGluThrTrpLeuGln.MetHisHisAlaThrLysGlnGluValVa 81
        ||:::::|||   :::|||    :::         ||||||   :::::
    134 ACATACAAACTACTGTCCAGGAAGAATAAAAACCAGAAGCAAGGTCTTCT 183

     81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
        ::::          |||||||||                    |||
    184 GAAA.........CCCCTGCAAAGT..................CCCACAA 206

     98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
                  |||:::|||
    207 AACCAGGCTGCTTAAGTACT..........................CCA 229

    115 ProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLe 131
          ::::::|||      ||||||   |||   :::|||:::      ::
    230 TCACACTCTCTAGCCCTTGAGCACACATGGATGAAGATCCTCAGGCACAT 279

    131 uHisLeuTyrGlnAspLeuLysSerArgGlyIle......LeuProAlaA 146
        :||||||:::   ||||||   |||   |||:::      |||
    280 TCACTTGCATCTTGATCTTCTCAGCCTTGGATTACTATGGCTTCTTCTAT 329

    146 snThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAla 162
          ::::::|||:::|||:::||||||:::   :::|||      |||:::
    330 TAGTCTCCATCCATAGTCTCTACCGGGGGCAGTCCCTCATCTGATGCACA 379

    163 Gly............GlyAlaAlaMetSerLysHisLeuThrAsnSer 174
        |||           |||      |||:::      |||:::::|||
    380 GGAGAGGTCCTTAATGGGCAGTTAATGACCCTGCTTCTGGCTGGGTCA 427
```

seq_name: gb_gss:AZ953717

seq_documentation_block:

```
LOCUS       AZ953717               632 bp    DNA     linear   GSS 27-APR-2001
DEFINITION  2M0219P02F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0219P02 F, DNA sequence.
ACCESSION   AZ953717
```

```
VERSION     AZ953717.1  GI:13824944
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0219  row: P  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0219P02"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      142 a    123 c    169 g    197 t      1 others
ORIGIN


alignment_scores:
            Quality:  103.50              Length:    170
              Ratio:  1.176                 Gaps:      9
 Percent Similarity:  51.765   Percent Identity: 22.941

alignment_block:
 US-09-674-779-2 x AZ953717/rev   ..

 Align seg 1/1  to reverse of: AZ953717  from: 1  to: 632

      42 LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSe 58
         |||::::::::::          |||||||||   |||   ::::::
     426 AAGCAACAACTNCCACATACACACACACACACACACACACACACA..... 382

      58 rHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisA 75
                            |||   :::   |||         :::::::
     381 ..................CACACAAAATGTTGGCCCACTTGCTATAACT 351

      75 laThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeu 91
         ::      ::::::|||:::      :::      |||   |||::::::
     350 GC......AATAATGTCATAACTCAGGAGCTTCCTCTGTAATCTCATATT 307

      92 GlyAsnTyrLeuProProMetSer...GlnLeuLeuThrThrAlaArgSe 107
               :::||||||      :::   ::::::::::::   :::|||::
     306 AAAATACACTTACCACAAAGGGAGCCAAAAGTGGTGGCCCTGTCAAGGCA 257

     107 rTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuT 124
         :|||:::|||                |||::::::|||
     256 ATGGGAAGCATGG............TATGACTTTCCAGAG.......... 229

     124 rpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
                  :::   :::||||||      :::   ::::::|||
     228 ..........CTACAGTCCCTTCATAGGCAGGATCATGTAAGATCA... 193

     141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl 157
            |||      |||::::::|||         |||      :::   ::
     192 GTGATATGCATGGCTTCAGCTCAG......TGTGTGGTGTCCAGTGAAAG 149

     157 uLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnS 174
         :   :::::::::   ::::::          |||||||||   |||:::
     148 CTGGCAAAGAGCACCATCAGCCCACAAAGAATCCAAGCATGGGACTCAC. 100

     174 erAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeu 190
                     |||   |||   |||:::::::::::::|||
      99 .............TGGGGTCCAAAGCTCCAGCTGCAAGCTCAGCTC... 67

     191 TyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAs 207
                          |||::::::|||||:::|||:::   ||
      66 .................TGGTGCCATCATTGGTTACAGCACTCCCCAAA 35

     207 nGlnAsnPhe 210
         |   ||||||
      34 CCCCAACTTT 25

seq_name: gb_gss:AZ175435

seq_documentation_block:
LOCUS       AZ175435               817 bp    DNA     linear   GSS 29-AUG-2000
DEFINITION  SP_0132_A1_H03_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=132 Col=5 Row=O, DNA sequence.
ACCESSION   AZ175435
VERSION     AZ175435.1  GI:8345803
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 132  row: O  column: 5
            Seq primer: T7
```

```
           Class: BAC ends
           High quality sequence stop: 817.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=132 Col=5 Row=O"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      181 a    198 c    211 g    227 t
ORIGIN


alignment_scores:
            Quality:   98.50             Length:    166
              Ratio:  1.119               Gaps:      6
 Percent Similarity:  53.012   Percent Identity: 24.096

alignment_block:
 US-09-674-779-2 x AZ175435/rev   ..

 Align seg 1/1  to reverse of: AZ175435  from: 1  to: 817

      14 SerSerMetLeuValAlaCysSerAlaProIleProThrAsnProGlnVa 30
         |||:::   |||   |||   :::      :::         |||   ||
     644 TCGGGTAAGCTACCCGCGATTGGACAAAAGTTACAATGCCCCCCGACGGT 595

      30 lSerProIleLysThrProSerValLeuIleThrLysAspLysIleGlyA 47
         |      |||   ||||||::::::       |||      :::|||
     594 CATTGCGATCCTAACACCAGGTATAAATAGCCACAAAACGGTGCTGGGCA 545

      47 spHisHisThrHisGluHis.............AspGluSerValSerHis 59
           |||   :::||||||:::           :::
     544 TCCATTCAAGTCATGAACAATTGTTACGGTCTGAACTTCGCTTCCGATTT 495

      60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHis........ 73
         :::   |||:::   :::   :::::::|||   :::   :::
     494 ATCTAACTTCATCAACAACATCATGTCTGGCAGAGGAAAGAGTGGAAAGG 445

      74 ....HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnS 89
             |||      |||:::   ::::::      :::|||   ::::::|
     444 CCCGCACCAAGGCAAAGACGCGCTCATCCCGTGCAGGGCTCCAGTTTCCA 395

      89 erArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAla 105
         ||        :::   :::   |||   :::|||||||         :::
     394 GTGGGACGTGTTCATCGGTTTCTCCGAAAGGGCAACTATGCAAAGAGGGT 345

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro.GluH 122
         |||   |||::::::|||   |||                 ||| |||:
     344 CGGCGGTGGAGCTCCTGTCTACAT.........GGCTGCCGTCCTAGAGT 304

     122 isLeuTrpGlyGlnIleValProThrLeu...HisLeuTyrGlnAspLeu 137
         ::|||   ::::::|||:::           :::      :::|||
     303 ACCTCACTGCCGAAATCTTGGAACTCGCAGGCAACGCTGCCCGCGACAAC 254

     138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
         |||:::     |||:::   ::::::   |||:::   :::|||   ::
     253 AAGAAATCTAGGATCATCACAAGCCACCTTCAACTC...GCTGTGCGTAA 207

     154 gAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLys 169
         ::::   |||||||||:::      ||||||:::::::::::::::
     206 TGATGAAGAACTCAACAAGCTTTTGGGTGGGGTGACGATCGCTCAA 161

seq_name: gb_est2:BM467307

seq_documentation_block:
LOCUS       BM467307               1270 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6471965 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5573098
            5', mRNA sequence.
ACCESSION   BM467307
VERSION     BM467307.1  GI:18516349
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1270)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12319  row: j  column: 11
            High quality sequence stop: 469.
FEATURES             Location/Qualifiers
     source          1. .1270
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5573098"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      272 a    475 c    242 g    281 t
ORIGIN


alignment_scores:
            Quality:   98.50             Length:    206
              Ratio:  1.048               Gaps:      6
 Percent Similarity:  45.631   Percent Identity: 24.757

alignment_block:
 US-09-674-779-2 x BM467307   ..

 Align seg 1/1  to: BM467307  from: 1  to: 1270

      13 IleSerSerMetLeuValAlaCysSerAlaProIleProThrAsnProGl 29
         ||||||||||:::      :::|||            ||||||:::|||
     744 ATCTCCTCCATAACCTCCACCTGCCTTCCCATCAGTCCTACCCACCCACC 793

      29 nValSerProIleLysThrProSerValLeuIleThrLysAspLysIleG 46
              |||:::|||:::|||                              |
     794 TCAACATCCAGTAAAATCTCCC.......................G 816

      46 lyAspHisHisThrHisGluHisAspGluSerValSerHisValGlyLeu 62
         ||   |||   :::|||
     817 GCCTACACATGTCTCAT.............................. 833

      63 GlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGl 79
                                                     |||
     834 .....................................ACACCAATTCC 844

      79 uValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr.Leu 95
          :::::::|||   :::   |||      |||:::::::::::::::: |||
     845 ACTCACCCGCTCCCGCCACTATTGTGCGTCCCAGATCTCCTCACATCCTA 894

      96 ProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGl 112
         ||||||      :::|||||||:::   :::::::::            ::
     895 CCGCCGGCTCTCGAACTTCTCGTGTTATCACAGCAACTGCCTCCTCACTC 944

     112 yHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValP 129
         ::::   |||:::   ||||||         |||   |||       :::|
```

```
     945 CCGCCTTCCTTTCCCTCTGCCCTCCCTCCCTCTTTGTGGAATCGCTACTC 994

     129 roThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla 145
         ||   ||||||      :::   |||:::::      |||
     995 CGCTCCTCCACCAGCTAAGTGCACTGTCAGCAGTGTGCATATCCGCGCGC 1044

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
             :::|||||||||   ||||||:::      |||
    1045 CCCCAACACATTCGCTCA...TATCGCACTGTACCCCTC..... 1080

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsn...SerAlaIleAspI 178
             ::::::::::::   |||      |||         |||:::
    1081 .......TCCTCAATCACACCACACACGCACAATCTCCCTATTATCTCCC 1123

     178 leTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGln 194
             :::|||:::          |||:::        :::   |||
    1124 CTCTTCTCCCATCCTGCATGTGGTAATCGCGTCCAAACGACTCGCCTCAA 1173

     195 AsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGl 211
         |||||||||                :::   ::::::      :::|||::
    1174 AATCGTCTCTACGCTCTCTCCCCCCGTGTTGCGCGCCCCACCCACTTCGC 1223

     211 yLeuGlyLeuTyrAla 216
         :   :::|||||||||
    1224 ATCATCTCTATACGCG 1239

seq_name: gb_gss:AG140443

seq_documentation_block:
LOCUS       AG140443                643 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-155D18.F, genomic survey sequence.
ACCESSION   AG140443
VERSION     AG140443.1  GI:16670121
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-155D18.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 643)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-155D18.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      143 a    240 c    101 g    158 t      1 others
ORIGIN
```

```
alignment_scores:
            Quality:   98.00            Length:    208
              Ratio:  0.990              Gaps:     13
 Percent Similarity:  47.596   Percent Identity: 26.923

alignment_block:
 US-09-674-779-2 x AG140443   ..

 Align seg 1/1  to: AG140443  from: 1  to: 643

      10 ThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProTh 26
         ||||||:::   :::::::::      :::         |||
      43 ACGACGTTCGCCCAGNCCTTGTATTGCTCTCCAGTACCTCCTTATCAAAG 92

      26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
          :::      :::|||||||||      ||||||            ||||
      93 ACACCTGCCTACTTCTCCCATCATGGAGCCGTCA.........CACAAAG 133

      43 spLysIleGlyAsp.His...HisThrHisGluHisAspGluSerValSe 58
         ||        ||| |||   |||||||||   |||           ::
     134 ATACACACACTGACACATAGACACACACACACAC............AC 171

      58 rHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisA 75
         :|||:::      ::::::|||   :::|||            |||::::
     172 ACACACACACACACACACACACACACACACACACACCCTGCCCCATGAAG 221

      75 laThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeu 91
         ::   ::::::|||:::|||      |||
     222 TTTACTCGAAAGAACTGGTTCCACTTCAA..................... 250

      92 GlyAsnTyrLeuPro.................ProMetSerGlnLeu.. 101
                 ||||||                 |||:::|||:::|||
     251 .........CTTCCTTTTCCTTAGTCTTGGTCACCTTTTTCTTCCTTGCC 291

     102 .LeuThrThrAlaArgSerTrpGln.....................AlaC 111
          |||      :::      |||:::                     :::|
     292 ACTGCAACCCTGCCTAGTGTGGCACCTGCTGAATGGCCATAAGTTCTGCT 341

     111 ysGlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIle 127
         ||:::|||                  |||   :::          |||
     342 GCTCACATCCACAGGTCCTGTCAGTCCCCACCAGG.........CAGCCT 382

     128 ValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuPr 144
         :::|||:::||||||   :::::::::|||::::::   |||:::|||||
     383 CTTCCCTCCTTGCACTA.TGGGAGTCCCTGAACACTCTCGGGCTGCTCCC 431

     144 oAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnC 161
         |            :::|||   |||   |||:::
     432 CATA.........CTGCGCATGGTGCTCCGAACG................ 456

     161 ysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp 177
                 ||||||   |||   ||||||
     457 ..........GCTGCCTCTTCTCTTCATCTC.................. 477

     178 IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr......Gl 192
         |||   |||::::::   |||:::      :::   |||      ::
     478 ...TGGTATCCATCCATTCCTATCTCCCCACCTACCACCTACTGGTCCTC 524

     192 uLeuGlnAsnArgLeuCysGln 199
         :::::::|||:::||||||:::
     525 ATTTTCAAACCATCTGTGTGAA 546

seq_name: gb_est2:BG214490

seq_documentation_block:
LOCUS       BG214490                621 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST34131 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG214490
VERSION     BG214490.1  GI:13740511
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 621)
  AUTHORS  Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
           Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
           , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
           ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
           ,J., Danzig,J. and Ducar,M.
  TITLE    Creation of genome-wide protein expression libraries using random
           activation of gene expression
  JOURNAL  Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE  21227151
COMMENT    Contact: Scott J. Cain
           Athersys, Inc.
           3201 Carnegie Ave, Cleveland, OH 44115, USA
           Tel: 216 431 9900
           Fax: 216 361 9596
           Email: scain@athersys.com
           High quality sequence stop: 495.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      210 a    160 c    125 g    126 t
ORIGIN


alignment_scores:
            Quality:   97.00            Length:    191
              Ratio:  1.115              Gaps:     12
 Percent Similarity:  45.550   Percent Identity: 26.702

alignment_block:
 US-09-674-779-2 x BG214490   ..

 Align seg 1/1  to: BG214490  from: 1  to: 621

      53 HisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTr 69
         |||:::||||||:::|||:::      :::   :::      |||   ||
      72 CACAGTGAGAGCCTATCAAACTAGTAAGTCACGAGTGACCATGAGCCCTG 121

      69 pLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaT 86
         |||||||:::         ::::::|||:::   :::      :::
     122 GCTTCAGCTC......CCAGTGGAACAGCAGCCAACCAGCCTGGAATACA 165

      86 yrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu 102
            |||   |||:::            ||||||         ||||||
     166 TACTCCTTTCCAAGAGCCAGCCTGCACTACCAGTCCCATGCCAGCTATAC 215

     103 ThrThrAlaArgSerTrpGlnAlaCys..................... 111
                    |||      |||
     216 CTACTG......TACTGGACACCCTGCTTCTAGCTCCACAGTGAGAGCCT 259

     112 .....................GlyHisGluProTyr...GlnLeuProP 120
                              |||||||||:::   |||||||||
     260 ATCAAACTAGTAAGTCACGAGTGACCATGAGCCCTGGCTTCAGCTCCCAG 309

     120 roGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAsp 136
         |||:::                  ||||||         |||
     310 TGGAGCAG..............CAGCCAACCAGCCTGGAATACATACTC 344

     137 Leu...LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerVa 152
         |||   ||||||:::      ||||||:::      |||:::      ::
     345 CTTTCCAAGAGCCAGCCTGCACTACCAGTCCCATGCCAGCTATACCTACT 394

     152 lTyr.................ArgAsnProGluLeuAsnGln....... 160
         :|||                 |||||||||:::::|||:::
     395 GTACTGGACACCCTGCTGTGAAAGAAATCCAAGAGTGAACAAGACCAAGT 444

     161 ..CysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIle 176
           |||:::   |||:::::::::         |||   |||:::   :::
     445 CTTGCTCTTATGGAACTTCTATT.........CTAGGGAATGGGGAAGTT 485

     177 AspIleTrpVal.ProAspLeuGluIleLysSerGlnAlaLeuTyrGluL 193
               ::::::    :::                          |||:::|
     486 GGTTATTATATAAAAAGAA.......................TATAAAT 511

     193 euGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsn 209
         |||||                        |||   ::::::::::|||
     512 TACAA.....................GAAGCAAACAAGAGAGAAAAT 537

     210 PheGlyLeuGlyLeuTyrAla 216
         :::      ||||||||||||
     538 TATACCACAGGGTTATATGCT 558

seq_name: gb_est2:BE273034

seq_documentation_block:
LOCUS       BE273034                764 bp    mRNA    linear   EST 13-JUL-2000
DEFINITION  601171358F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3544670 5',
            mRNA sequence.
ACCESSION   BE273034
VERSION     BE273034.1  GI:9147389
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 764)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM240  row: d  column: 15
            High quality sequence stop: 612.
FEATURES             Location/Qualifiers
     source          1. .764
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3544670"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      165 a    218 c    204 g    177 t
ORIGIN


alignment_scores:
            Quality:   97.00            Length:    257
              Ratio:  0.858              Gaps:     14
 Percent Similarity:  43.969   Percent Identity: 21.401

alignment_block:
```

```
US-09-674-779-2 x BE273034/rev   ..

Align seg 1/1  to reverse of: BE273034  from: 1  to: 764

      9 IleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIlePr 25
        :::   |||:::   :::::::::::::::::::|||      |||
    740 GTGTGGACCATCTGTAAAGCGTTAGTTATTAGTTGC......CCCAATTC 697

     25 oThrAsnProGlnValSerProIleLysThrProSerValLeuIleThrL 42
         :::|||||||         ||||||   |||:::|||
    696 CGGTTCACCGCAA.........ATTAAAAGGCCCCAGGTC.......... 666

     42 ysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal... 57
                             :::||||||||||   :::   :::
    665 ......................AACCATGAGCACCATCAGATCCTCCAG 639

     58 ..................SerHisValGlyLeuGlnAlaHisPheGluTh 68
                          :::   |||||||||:::          ||
    638 GTGTTTTGGTTGTGAGATGGATTTGTAGGGCTCAACTACAGCTGTGGAAC 589

     68 rTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnA 85
        ||||                                          :
    588 CTGG....................................GGTGCTG 578

     85 laTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeu 101
        ::   :::::::::         :::::||||||      ::::::
    577 GGTAAATGGAGAACTCCAGCTTGGACTTCTTGCCATAATCAACTGAGAGA 528

    102 LeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLe 118
           :::   :::|||:::                  :::|||      ::
    527 CGTTCCATGAGCAGGGAG..............GTGAACCCAGAACCAGT 493

    118 uProProGluHisLeuTrp.....................GlyGlnI 127
        :||||||      ||||||                      |||   :
    492 TCCCCCACCAAAGCTGTGGAAAACCCAAGAAGCCCTGAAGACCGGTGCAC 443

    127 leValProThrLeu..............HisLeuTyrGlnAspLeuLys 138
        ::|||      |||            |||      :::||||||
    442 TGGTCAGCCAGCTTGCGAATTCGGTCCAACACAAGGTCAATGATCTCCTT 393

    139 SerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAs 155
        ::::::|||:::::::|||::::::      |||   |||::::::
    392 GCCAATGGTGTAGTGCCCTCGGGCATAGTTATTGGCAGCATCTTCCTT.. 345

    155 nProGluLeuAsnGln.CysAlaGlyGlyAlaAlaMetSerLysHisLeu 171
                 :::::: |||:::|||      :::
    344 ....GCCTGTGATGAGCTGCTCAGGGTGGAAGAGCTGGCGGTAGGTGCCA 299

    172 ThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGl 188
        ::::::::::::      :::|||||||||:::|||:::   :::
    298 GTGCGAACTTCATCATGACTGTGGGTTCCAAGTCTACAAACACAGCCCGG 249

    188 nAlaLeuTyrGluLeuGlnAsnArgLeuCysGln.............. 199
         |||                 |||   ||||||
    248 GGCA..................CGTGCTTGCCAGCGCCCGTCTCACTGAA 217

    200 .................TyrTrpLeuGluHisGly...GluAsnGlnAsn 209
                        :::   |||   ||||||      :::   :::
    216 GAAGGTGTTGAAGGAGTCATCTCCTCCCCCATGGTCTTGTCACTTGGCAT 167

    210 PheGlyLeuGlyLeuTyrAla.......................ThrGl 218
        ::::::::::||||||   |||                     :::||
    166 CTGGCCATCGGGCTGGATGCCGTGTTCCAGGCAGTAGAGCTCCCAGCAGG 117

    218 yAlaIleHisLeuAspThr 224
        |      :::||||||||||
    116 CATTGCCAATCTGGACACC 98

seq_name: gb_est2:BI454968

seq_documentation_block:
LOCUS       BI454968               874 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603173541F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5252650 5',
            mRNA sequence.
ACCESSION   BI454968
VERSION     BI454968.1  GI:15245624
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11638  row: b  column: 11
            High quality sequence start: 59
            High quality sequence stop: 395.
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Mus musculus"
                     /strain="C57/B6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5252650"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      265 a    286 c    169 g    154 t
ORIGIN


alignment_scores:
           Quality:   97.00            Length:    269
             Ratio:  0.843              Gaps:     15
 Percent Similarity:  42.751   Percent Identity: 23.420

alignment_block:
 US-09-674-779-2 x BI454968   ..

 Align seg 1/1  to: BI454968  from: 1  to: 874

      8 PheIleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIl 24
        |||         ::::::   |||   ::::::||||||:::|||   ::
    234 TTTACCATGGACTTCTTAATGAGTGGTGTAGCAGCCTGTGGAGCCTGTGT 283

     24 eProThrAsnProGlnValSerProIleLysThrProSerValLeuIleT 41
        :   |||||||||   ::::::   :::||||||         |||
    284 GTTCACCAATCCC...TTGGAGGTTGTGAAGACCAGAATGCAATTGCAGG 330

     41 hrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57
          :::               ||||||||              |||:::
    331 GCAGAACTGCAGGCCCCAGGCACCATACCAGCGCCACTACCGAAACGTCA 380

     58 SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMet..... 72
          |||:::   ||||||   |||:::::::   |||:::   :::
    381 TTCCACGCCATTCTTCAA...CATTGGCAAGGATGGATGGCACTAAGCAT 427

     73 .......His.HisAlaThrLysGlnGluValValArgTyrGlnAlaTyr 86
              ||| :::|||   :::   :::::::::   |||      |||
    428 GCCCTGACACAAGGGCCTGGAGACCAGACGCCATCATGTACCCCAGATAC 477
```

```
     87 .LeuGlnSer......ArgLeuGlyAsnTyrLeuProProMetSerGlnL 101
        |||::::::      ||||||||
    478 ATTGATGAATGGCATCCGACTGGGA........................ 502

    101 euLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
                          |||:::   |||     |||
    503 ......................CAAACTACCGGGACTGGCCCG...... 523

    118 LeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134

    523 ................................................. 523

    134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
              :::|||||   |||   |||   :::|||
    524 ..........AGATCTCGGAGGATATACCCAGACCACACT.......... 553

    151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMet 167
                               ||||||   |||:::
    554 ........................CAATGCAACGGCAACCCCAACACA 577

    168 SerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro...AspLe 183
        :::                 :::|||   :::   |||:::|||   |||||
    578 GACACATGATCACCAGCAACAGCAACGCAAGCATGGACGCCACATGACCT 627

    183 uGlu.................IleLysSerGlnAlaLeuTyrGluLeuG 194
        ||||                |||   |||::::::      :::   :
    628 CGAGACGTCCAATGCCGCACCAATAATCTCCGAGACACACACCCAAACAA 677

    194 lnAsn...ArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsn 209
        ::|||   |||      ::::::      :::|||:::::::::::::::
    678 AAAATGGACGAAGAACAAGACATACAAGGAAACACGCACACGACCACACG 727

    210 PheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHisLeu.......... 222
           |||   :::|||      |||::::::|||||
    728 ...GGAAACTCGCTGCAACAGACGAACACGATACACACCCCACTACAAAC 774

    223 .....................AspThrGlnGlyPheArgLysTrpGlyA 232
                            ||||||   |||   |||
    775 CACACCCATCCCCCCACAACCAGACACACCAGGCAACAGG.......... 814

    232 laGlnPheSerGluThrAsnSerIleCysArgHisValLeuProLysAsn 248
                     |||:::::::   :::|||||||:::   |||   :::
    815 ..............ACGCATACCACAGCACGACACGCGCACCCGCCAAAA 850

    249 Lys 249
        :::
    851 CGA 853
```

seq_name: gb_gss:CNS04OFM

seq_documentation_block:

```
LOCUS       CNS04OFM                813 bp    DNA     linear   GSS 21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            124008 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL299947
VERSION     AL299947.1  GI:8039088
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 813)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 813)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="124008"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG124BH04SP1~end :
                     PUC-Ori"
BASE COUNT      234 a    191 c    189 g    189 t     10 others
ORIGIN
```

```
alignment_scores:
           Quality:   96.00              Length:     61
             Ratio:  2.462                Gaps:      0
 Percent Similarity:  63.934   Percent Identity: 34.426

alignment_block:
 US-09-674-779-2 x CNS04OFM/rev   ..

 Align seg 1/1  to reverse of: CNS04OFM  from: 1  to: 813

     130 ThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAs 146
         |||   :::|||   |||:::|||            :::||||||   ::
     209 ACCTCTGAGCTCATCCAGGAGCTGATTTACTCTGAGATGCTGCCAGAGTC 160

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
         :::::::   |||            |||||||||:::         :::|
     159 CTCCTCTCAGCGGGTCTGTAGTTCAAACCCAGAGATGCTGCCTGGCTCGG 110

     163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
         ||   :::::::::   :::|||      |||   :::|||::::::|||
     109 GCACCTCTGGGCTCCCACAACACCAGGAGAATGTGAGTATCTCCCTCTGG 60

     180 ValProAspLeuGluIleLysSerGlnAlaLeu 190
         |||||||||||||||:::   :::   :::|||
      59 GTGCCAGACCTGGAAGTCTGGGATGCATCACTT 27
```

seq_name: gb_est2:BE442326

seq_documentation_block:

```
LOCUS       BE442326                454 bp    mRNA    linear   EST 25-JUL-2000
DEFINITION  925017E08.x1 C. reinhardtii CC-2290, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BE442326
VERSION     BE442326.1  GI:9441841
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
```

```
           Duke University
           Durham, NC 27708-1000, USA
           Tel: 919 613 8164
           Fax: 919 613 8177
           Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-2290 wild type mt- S1 D2"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-2290, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library was constructed by John Davies and
                     Jeffrey McDermott. RNA was isolated from strain CC-2290
                     (Minnesota isolate of C. reinhardtii) grown to mid-log
                     phase in TAP (acetate containing) medium in the light.
                     PolyA mRNA was purified, and cDNA was synthesized and
                     directionally cloned into lambda ZAP II (Stratagene) in
                     the EcoRI (5') and XhoI (3') sites. pBluescript II SK-
                     plasmids were excised from the lambda ZAP clones by
                     superinfection with ExAssist (Stratagene) phage. The
                     library was normalized using method 4 described in Bonaldo
                     et al (1996) Genome Research 6: 791-806."
BASE COUNT      143 a    140 c    101 g     70 t
ORIGIN


alignment_scores:
            Quality:   95.50            Length:     84
              Ratio:  1.910              Gaps:      5
 Percent Similarity:  59.524   Percent Identity: 35.714

alignment_block:
 US-09-674-779-2 x BE442326   ..

 Align seg 1/1  to: BE442326  from: 1  to: 454

      60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
         :::|||   ||||||||||||:::|||     |||:::::::   |||::
     183 ATTGGCCAACAAGCACATTTCGACACA......CAGCAAAACGCAGCAGC 226

      76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
         :||||||::::::::|||   ::::::        :::|||   |||:
     227 AAAACAACACGCTATCCGCCAAGACACA......TAAAATCGCGCTGGCA 270

      93 snTyrLeuProProMetSerGlnLeuLeuThrThrAlaArg...SerTrp 108
         ::             |||   ||||||      |||   ||||||
     271 CC......AATCAAAGGCACCAAAAACTCACAAAAAGACGGCAAAGCTGG 314

     109 GlnAlaCysGly......HisGluProTyrGlnLeuProProGluHisLe 123
         |||   |||:::      |||:::|||   |||      |||:::
     315 CAAAACTGCGCTCAAGTTCACAGGCCAGCGCAGGCGTGCCCACATGTGCG 364

     123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
          ||||||           :::   |||:::|||:::   :::   :::|
     365 CTGGGGCACGCGCTATCGGAACAAGCACGTGTACCGTGCTGTCGCCGACA 414

     140 rg 140
         ||
     415 GA 416

seq_name: gb_gss:BH032955

seq_documentation_block:
LOCUS       BH032955               774 bp    DNA     linear   GSS 17-JUL-2001
DEFINITION  RPCI-24-294G15.TV RPCI-24 Mus musculus genomic clone RPCI-24-294G15
            , DNA sequence.
ACCESSION   BH032955
VERSION     BH032955.1  GI:14804300
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-294G15.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 294  row: G  column: 15
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-294G15"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      229 a    265 c     85 g    195 t
ORIGIN


alignment_scores:
            Quality:   95.00            Length:     76
              Ratio:  2.159              Gaps:      1
 Percent Similarity:  57.895   Percent Identity: 27.632

alignment_block:
 US-09-674-779-2 x BH032955   ..

 Align seg 1/1  to: BH032955  from: 1  to: 774

       3 AsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuValAl 19
         |||   |||      :::|||||||||   :::   ::::::::::::
      83 AACACCAACATCACCATCATCACCACCACTGTTATCGCAATCATCATCAT 132

      19 aCysSerAlaProIleProThrAsnPro...........GlnValSerP 32
         ::::::         |||:::   |||            :::|||
     133 AGCCGCAATTGTCACACCTTCACCCCCAGTCATCACAAGTACCATTAGCA 182

      32 roIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHis 48
         :::   |||   |||::::::   |||:::|||:::      :::|||
     183 CTGTGGTCACCATTTCCATCATCACAACCAGGGACAACCACCATCAACAT 232

      49 HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHi 65
         |||   |||:::|||   :::      :::|||      |||:::   ||
     233 CATCACCACCAACACTATCACCATCACCAACACCATCACCTCCATCATCA 282

      65 sPheGluThrTrpLeuGlnMetHisHis 74
         |::::::         :::   ||||||
     283 TTATAATCATCAACATCACCACCACCAC 310

seq_name: gb_est2:BF627823
```

```
seq_documentation_block:
LOCUS       BF627823               808 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEb0005009f Hordeum vulgare seedling shoot EST library
            HVcDNA0002 (Dehydration stress) Hordeum vulgare cDNA clone
            HVSMEb0005009f, mRNA sequence.
ACCESSION   BF627823
VERSION     BF627823.2  GI:13090019
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex drought-stressed seedling shoot cDNA
            library
  JOURNAL   Unpublished (2001)
COMMENT     On Dec 19, 2000 this sequence version replaced gi:11891981.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 275
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 4
            High quality sequence stop: 753.
FEATURES             Location/Qualifiers
     source          1. .808
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEb0005009f"
                     /clone_lib="Hordeum vulgare seedling shoot EST library
                     HVcDNA0002 (Dehydration stress)"
                     /tissue_type="Seedling shoot"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Seeds were surface sterilized then germinated under axenic
                     conditions in the dark at room temperature on filter paper
                     with water, nystatin and cefotaxime in covered
                     crystallization dishes. Five-day old seedlings were
                     incubated at 90% RH for 24 hr. Shoots were then harvested,
                     total RNA was prepared, poly(A) RNA was purified, one
                     primary unamplified cDNA library was made, 600000 pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids.
                     These steps were performed in the TJ Close laboratory at
                     the University of California, Riverside (Choi, Close,
                     Fenton). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      101 a    256 c    321 g    130 t
ORIGIN
```

```
alignment_scores:
           Quality:   95.00            Length:    177
             Ratio:  1.159              Gaps:      7
 Percent Similarity:  46.328   Percent Identity: 22.034

alignment_block:
 US-09-674-779-2 x BF627823/rev   ..

 Align seg 1/1  to reverse of: BF627823  from: 1  to: 808

      19 AlaCysSerAlaProIleProThrAsnProGlnValSerProIleLysTh 35
         ||||||::::::|||   ||||||   ||||||:::||||||      ::
     522 GCCTGCGCCAGCCCCTACCCTACC...CCCCAGGCCAGCCCCCGCGGCAG 476

      35 rProSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHisG 52
         :|||:::   |||
     475 CCCCACCCCCCTC..................................... 463

      52 luHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThr 68
                                        |||:::::::|||   :::
     462 .........................CGGCTTCGGTGCCATTCCCGTCCC 439

      69 TrpLeuGlnMetHis.............................HisAl 75
            |||::::::|||                              |||
     438 TCCCTGCGCCAGCACGCGGGCAGGGCAGGGCAGGGCCCCAGCCGGCACCT 389

      75 aThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuG 92
               :::::::::::|||:::|||   |||      ::::::   :
     388 CCGCGCACGCCAGGCGCTGCGCTTCCAGTAGTACTCCCCCCAGCACCACA 339

      92 ly..................AsnTyrLeuProProMetSerGlnLeuLeu 102
         ::                  :::      |||||||||:::
     338 GCTAGCCGGTGGCCTAGTACAGCAGGTAGCCCCCCATGCAG......... 298

     103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
               ::::::::::|||::::::||||||         :::|||   ||
     297 TTGTGCAGCCAGGAGTGGAGCGGGTGTGGCAGCGCCCGCCACCAGCCGCC 248

     119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136
         ||||                 :::      :::   |||||||||:
     247 GCCGCCGGGGACG......ACCGCGCTGAGGAAGATCTCGTTGTACCAGT 204

     136 spLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerVal 152
         ::::::::||||||
     203 CCGTCTCCTCCCGG..................................... 190

     153 TyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLy 169
         ::::::|||:::   ::::::   |||::::::||||||   :::
     189 ...AACAGGCCCCAGTAATCCTCCCCCGCCGCCGCCGCCGCCGCCGCGC 143

     169 sHisLeuThrAsnSerAlaIleAspIleTrp 179
            :::::::::|||:::         |||
     142 CGCCATTGCCGGCAGCGGTGGGTGTGCCTGG 112

seq_name: gb_est2:BE974669

seq_documentation_block:
LOCUS       BE974669               1044 bp    mRNA    linear   EST 04-OCT-2000
DEFINITION  601680707R2 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:3951145 3',
            mRNA sequence.
ACCESSION   BE974669
VERSION     BE974669.1  GI:10588005
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1044)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM818  row: m  column: 02
           High quality sequence stop: 11.
FEATURES             Location/Qualifiers
     source          1. .1044
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3951145"
                     /clone_lib="NIH_MGC_83"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.4 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      214 a    231 c    331 g    263 t      5 others
ORIGIN


alignment_scores:
            Quality:   95.00            Length:    196
              Ratio:  0.990              Gaps:     15
 Percent Similarity:  48.980   Percent Identity: 27.551

alignment_block:
 US-09-674-779-2 x BE974669/rev   ..

 Align seg 1/1  to reverse of: BE974669  from: 1  to: 1044

      23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
         |||   |||   ||||||:::|||:::   ::::::   |||   :::
     919 CCTTCTCCCCAGAACCCAGACGTAACAAGGGTTCGCCAACCCAGGCTCCA 870

      39 uIleThrLys.AspLysIleGlyAspHisHisThrHisGluHisAspGlu 55
          |||       |||   :::|||      |||      ||||||
     869 GATA...CCACGACCCCGTTGGGCCCCCCCACCACAAAGAGCATTTTTAT 823

      56 .SerValSerHisValGlyLeuGlnAlaHisPheGluThrTrp...LeuG 71
         ::::::      ||||||   |||            :::|||   :::
     822 TAACCTCCACAGCGTCGGAACACAA...........TCATGGGGTATTG 785

      71 lnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeu 87
           ||||||   ::::::   :::
     784 GCGACCACCACCCCGCCCGCCTCAAG..................... 759

      88 GlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrTh 104
         |||      :::|||       ||||||   :::        :::::
     758 CAACCATGGGTGGGC.........CCCCCACGAACCCCATCCTCCTCAAA 718

     104 rAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProG 121
         :||||||||||::::::   :::   :::      ::::::||||||||
     717 TGCTCGCTCGTATCACCACGCGAAGAGATGGTTCTTCTCTTTGCCACCAG 668

     121 luHis.LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLe 137
           ||| ||||||:::                 |||||||||:::
     667 GCCACCTTGTGGTCG...........TGTTGGCACCTATATCGCGGGCA 630

     137 uLysSerArg......GlyIleLeuProAlaAsnThrGlnIleArgServ 152
         ::::::      :::      |||   ||||||:::   ::::::
     629 AACAAATCAATGGGGCGCACACCGGCCCCCTAATACAGACAAGAAGGCCT 580

     152 alTyrArgAsnProGluLeuAsn...GlnCysAlaGlyGlyAlaAlaMet 167
           :::|||:::|||      :::   |||         ||||||
     579 CCTGGCGTGGTCCTGCGCGCCAATAGCAACCACCCCGCGGCGCGCTCCGA 530

     168 SerLysHis.........LeuThrAsnSer..................... 174
         ::::::|||           |||||||||
     529 GGGCGCCACCCGTCGGTCAAAACCAACTCGTGATAATCAGGGGTCGCACN 480

     175 .AlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeuT 191
          ||||||:::                  :::   ::::|||
     479 CAGAATAGATGTT.................GTGCACACACAACAGAGAG 448

     191 yrGluLeuGlnAsnArg......LeuCysGln 199
         :::|||::::::|||      |||||||||
     447 AGCACCTTATGAGTCGAGAAGATTTATGTCAA 416

seq_name: gb_est2:BI728822

seq_documentation_block:
LOCUS       BI728822               599 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  1031101G02.y1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI728822
VERSION     BI728822.1  GI:15704517
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      114 a    186 c    204 g     95 t
ORIGIN


alignment_scores:
            Quality:   94.50            Length:    184
              Ratio:  1.062              Gaps:     10
```

```
 Percent Similarity:  48.370   Percent Identity: 25.543

alignment_block:
 US-09-674-779-2 x BI728822/rev   ..

 Align seg 1/1  to reverse of: BI728822  from: 1  to: 599

       8 PheIleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIl 24
         :::   |||:::   |||:::||||||||||:::|||
     595 TACTCCACCTCCTGGATCACGTCCATGTTGGCGGCA..............  560

      24 eProThrAsnProGlnValSer.ProIleLysThrProSer.ValLeuIl 40
          |||::::::|||   :::::: |||:::      |||::: :::|||
     559 .CCTGCTCGGCCCACGTTGACCTCCACTGGGTCCACCAGCGCACTCTCCG 511

      40 eThrLysAspLysIleGlyAspHisHisThrHisGluHisAsp....... 54
            :::     :::|||   |||         ||||||:::
     510 CAAACGCCTTGATCTTGGCGGGCATGGTCGCGCTGAACATGAGTGTCTGC 461

      55 ...................GluSerValSerHisValGlyLeuGlnAla 64
                            ::::::|||:::||||||   |||:::|||
     460 CGCTGCCCCCTGAAAAAGGACAGAATGTCGCGCACGTCGTCCTCAAAGCC 411

      65 HisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValVa 81
         |||:::               :::||||||||||            |||
     410 CACCTG..............GTCCACCATGCG............GTCTG 388

      81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProM 98
            |||:::   :::   :::   :::      :::         :
     387 CCTCGTCCAGACACAGGTAGCGGCAGATGTCCAGGTTCATGCGCTTCTTG 338

      98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
         :::::|||:::|||   :::|||          :::||||||||||:::
     337 TGCAGCAGGTCCTTGAGGCGGCC............GGGTGTGGCCACCAC 300

     115 ProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLe 131
            |||   ||||||:::::::|||
     299 GGCGTGCAGCCCCCGCCGCAGCGTCTC..................... 273

     131 uHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrG 148
                      :::   :::   |||:::      |||::::::|
     272 ..............GTACATGGGCTTGGGGTCGATGCCGCCGATGCAC 239

     148 ln............IleArgSerValTyrArgAsnProGluLeu...... 158
         ||           :::|||   :::   |||      :::|||
     238 AGCATCACCCGCACCTCCGGGTAGCGGTCCGCGTTCAGCGCCTGGCAGTA 189

     159 ........................AsnGlnCysAlaGlyGlyAla 165
                                 :::|||  ||||||||||||
     188 CATGTCGGCGATCTCGTACGTCTGTGTGGCCAGCTCGCGGGAGGGGCA 141

seq_name: gb_est2:BE402495

seq_documentation_block:
LOCUS       BE402495               605 bp    mRNA    linear   EST 21-JUL-2000
DEFINITION  CSB008E09F990908 ITEC CSB Wheat Endosperm Library Triticum aestivum
            cDNA clone CSB008E09, mRNA sequence.
ACCESSION   BE402495
VERSION     BE402495.1  GI:9361963
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Appels R
            Div. of Plant Industry, CSIRO
            Canberra ACT 2601 AUSTRALIA
            Tel: 61 62 465496
            Fax: 61 62 465000
            Email: rudi@pi.csiro.au
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="Triticum aestivum"
                     /cultivar="Wyuna"
                     /db_xref="taxon:4565"
                     /clone="CSB008E09"
                     /clone_lib="ITEC CSB Wheat Endosperm Library"
                     /tissue_type="endosperm"
                     /dev_stage="8-12 days post anthesis"
                     /lab_host="Escherichia coli SOLR"
                     /note="Vector: Lambda Zap/Bluescript; Site_1: XhoI;
                     Site_2: EcoRI; Plants grown in Phytotron with 18C/13C
                     (day/night) 16 hour light. M13 Reverse sequencing primer
                     used.  1.0 Kbp average insert size."
BASE COUNT      271 a    260 c     39 g     35 t
ORIGIN


alignment_scores:
           Quality:   94.50             Length:    166
             Ratio:  1.212               Gaps:      8
 Percent Similarity:  46.988   Percent Identity: 24.699

alignment_block:
 US-09-674-779-2 x BE402495   ..

 Align seg 1/1  to: BE402495  from: 1  to: 605

      26 ThrAsnProGlnValSerProIleLysThrProSerValLeuIleThrLy 42
         |||:::|||:::   ||||||   ::::::|||:::            ||
     128 ACCACACCCCACAAATCACCACACAACAACCCCAACCCGACACACCACAA 177

      42 sAspLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerH 59
         |:::|||           ||||||::::::|||      |||::::::|
     178 ACAAAAAACCAAACACAAACACACACAACAACACACCACATCCATAACAC 227

      59 isValGlyLeuGlnAlaHisPheGluThrTrpLeu.............. 70
         ||            :::|||      |||   |||
     228 AC.........CCAACACACAACCCAACAACACTACACAACAACCTCCAC 268

      71 ........................GlnMetHisHisAlaThrLysGl 78
                                 :::   ||||||   |||:::||
     269 AACACACACGAAGAACCACAAAGCACTCACAACCACCACATCACAAACCA 318

      78 nGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrL 95
         |::::::::::::
     319 ACAAGCAATCAAACA................................ 333

      95 euProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys 111
           |||      :::|||      ||||||
     334 ..CCACACAAGCAACAAAACAAAACAACCCAA.................. 363

     112 GlyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleVa 128
           ::::::|||:::|||      |||:::            |||
     364 ...AACCAACCACACCAGACGACACCCCAACCCACAACCCACCAA..... 405

     128 lProThrLeuHis.LeuTyrGlnAspLeuLysSerArg............ 140
          |||||||||||| ::::::|||:::   |||::::::
     406 .CCGACACTGCACCATCCACCAATCCACCAAACAACAACCAAAAAATACA 454

     141 .........GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
                    :::      |||:::|||::::::      :::   |||
     455 AACAAAACATCCCCCACACCATCAAACAATAACACCCCGACCAACTACAG 504
```

```
     154 gAsnPro......GluLeuAsnGlnCysAlaGlyGlyAlaAlaMet 167
     505 CAATCCCCCAGAACAGCCAACCCCATGCCTGGGGAGCGCCCCGCTA 550

seq_name: gb_est2:BG845267

seq_documentation_block:
LOCUS       BG845267                700 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  1024009E05.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG845267
VERSION     BG845267.1  GI:14226451
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      143 a    253 c    181 g    122 t      1 others
ORIGIN


alignment_scores:
            Quality:   93.50            Length:    227
              Ratio:  1.016              Gaps:      9
 Percent Similarity:  40.529   Percent Identity: 23.348

alignment_block:
 US-09-674-779-2 x BG845267   ..

 Align seg 1/1  to: BG845267  from: 1  to: 700

      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
     155 ACGCTTCTCCTCTCGTGCCTGGTCCGCGCCCTTTGCGCCTGCCCTTACTT 204

      27 nProGlnValSerProIle.........LysThrProSerValLeuIleT 41
     205 GCCTCAGGTTTCACCACTCACCCACCAACACACCAGTAGCGTCCACATCA 254

      41 hrLysAspLys...IleGlyAspHisHisThrHisGluHisAspGluSer 56
     255 TCAACACTAAACCCATCAGTTGTCACCACTCCAAAATCTGCCAACCCAGC 304

      57 ValSerHisValGlyLeu..................GlnAlaHisPheGl 67
     305 ATGGAAGGCGGCGGCACCCGGGTTCTGGCCAGCGACAGAGCCCATGCACA 354

      67 uThrTrpLeuGlnMetHis.............................. 73
     355 GACCACGATTGCCATCCACACCACGCTGCGTTGCGCAAAGTAAAGGCCAG 404

      73 ..................................................... 73
     405 CCGCCGGCGACTGCGAGCAGGACACATCAGCCTCTCCCAGCCGCCAAGGC 454

      74 ...HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSe 89
     455 GTACATGGCGCCAAACAACCCGACGTGCGC...CAGAGCCACATTCGGGG 501

      89 rArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaA 106
     502 GCTCCCGCGTCGGGGGCCGCCGCCAATG......CTTTTTAACACACCGC 545

     106 rgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHis 122
     546 GGTCT.................................................. 550

     123 LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSe 139
     551 .....................TCAGTGTACAGGTATCAGGACTGGACGCT 579

     139 rArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnP 156
     580 GCGCGCCATGTCGCCTGTCAGCAAGCAACACGTCAGC...CAGCGCCGGT 626

     156 roGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
     627 CTTGGGCCGGCCGAGCAAACGGCGGC.......................... 652

     173 AsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAl 189
     653 .....................................................CAGGC 657

     189 aLeuTyrGluLeuGlnAsnArgLeuCysGln 199
     658 CATTTACAGGNCGGCCAGGTCGTTCTGCCAG 688

seq_name: gb_est2:BG840148

seq_documentation_block:
LOCUS       BG840148                550 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  MEST8-E10.T3 ISUM3-TL Zea mays cDNA clone MEST8-E10 3', mRNA
            sequence.
ACCESSION   BG840148
VERSION     BG840148.2  GI:14242972
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Qiu,F., Cui,F., Guo,L., Ashlock,D.A, Wen,T.J. and Schnable,P.S.
  TITLE     Expressed Sequence Tags from B73 Maize Seedlings and Silks
  JOURNAL   Unpublished (2001)
COMMENT     On May 25, 2001 this sequence version replaced gi:14206470.
            Contact: Patrick S. Schnable
            Schnable Laboratory
            Iowa State University
            G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
            Tel: 515-294-0975
```

```
           Fax: 515-294-2299
           Email: schnable@iastate.edu
           PCR PRimers
           FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)
           BACKWARD: T3 (ATT AAC CCT CAC TAA AG)
           Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="MEST8-E10"
                     /clone_lib="ISUM3-TL"
                     /tissue_type="Seedling and silk"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3PAC; Site_1: EcoRI; Site_2: NotI;
                     ds-cDNA molecules were generated as follows. First-strand
                     cDNA was prepared from oligo-dT selected mRNA by priming
                     with a NotI  oligo-dT primer (5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT).  The
                     resulting DNA:RNA hybrid was treated with RNase H and used
                     as a  template for DNA PolI-catalyzed second strand
                     synthesis. After the  addition of EcoRI adaptors, the
                     ds-cDNAs were digested with NotI and  size-selected. The
                     resulting molecules were directionally cloned  into the
                     EcoRI and NotI sites of the pT7T3PAC vector."
BASE COUNT       72 a    110 c    188 g    179 t      1 others
ORIGIN


alignment_scores:
            Quality:   93.00             Length:    184
              Ratio:  1.134               Gaps:      7
 Percent Similarity:  44.565   Percent Identity: 21.739

alignment_block:
 US-09-674-779-2 x BG840148/rev   ..

 Align seg 1/1  to reverse of: BG840148  from: 1  to: 550

      22 AlaProIleProThrAsnProGlnValSerProIleLysThrProSerVa 38
         ||||||   |||   :::|||:::   :::|||   :::   |||:::::
     523 GCCCCCAAGCCAAAGGAGCCAGAACAAGAACCGCCACAGCTACCACAAGC 474

      38 lLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspG 55
         :
     473 A..................................................... 473

      55 luSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGln 71
                                            :::      ||||||
     472 ................................CAACCCCAACTTCAG 458

      72 MetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGl 88
         :::   |||:::      |||:::   ::::::   |||:::   |||::
     457 CTG...CATACTGAACCACAGCAACCCATGCAATCTCAAGTTCAGTTGCA 411

      88 nSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrA 105
         :|||:::      :::||||||   |||:::   ||||||
     410 TTCGCAAACACAACACTACCTCCAACCTCAACTGCAGCTC.......... 371

     105 laArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGlu 121
                               |||:::|||   |||||||||
     370 ...................CACCATCAGCCACAACAGCTCCCACAAGTG 341

     122 ......HisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAs 136
               |||         |||:::   |||   :::|||||||:::   ::
     340 CAGCTGCACTCTCAGCCGCAGCTGCAACCCCAAGTGCATCTGCACCCTCA 291

     136 pLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg....... 150
         :   :::           ||||||   :::   |||::::::
     290 GCCGCAG............TTGCCCCCACAATTGCAGGTCCATCAGCAGC 253

     151 ...................SerValTyrArgAsnProGluLeuAsnGln 160
                                  :::|||:::::::::||||||::::::
     252 TGCAGCAACCCCCACAAGTGCAGGTCCACCAGCAGCCCGAAGTGCAGCCC 203

     161 CysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAs 177
               :::             |||   :::   |||::::::
     202 CAAGAAGCGCAGTTGCAATCATCAGCCCAGCAAACCTCCCAACCG..... 158

     177 pIleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuG 194
                 |||:::   :::::::::||||||:::      :::   |
     157 .......CAACCTCAGGCACAACTACAATCACAGGGGCACTCACAAGCGC 115

     194 ln 194
         ||
     114 AG 113

seq_name: gb_est2:BF578004

seq_documentation_block:
LOCUS        BF578004               1939 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION   602091838F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4206464 5',
             mRNA sequence.
ACCESSION    BF578004
VERSION      BF578004.1  GI:11651716
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 1939)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9767  row: k  column: 09
             High quality sequence start: 7
             High quality sequence stop: 126.
FEATURES             Location/Qualifiers
     source          1. .1939
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4206464"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      716 a    623 c    424 g    176 t
ORIGIN


alignment_scores:
            Quality:   92.50             Length:    192
              Ratio:  1.005               Gaps:     10
 Percent Similarity:  47.917   Percent Identity: 23.958

alignment_block:
 US-09-674-779-2 x BF578004   ..

 Align seg 1/1  to: BF578004  from: 1  to: 1939

      24 IleProThrAsnProGlnValSerProIleLysThrPro.......... 36
         :::|||::::::||||||      |||:::
     219 CTTCCATCAACCCCACAGAATCCACCTGTCATGCTATCCTGGATGTGCCC 268
```

```
 37 .....SerValLeuIleThrLysAspLysIleG 46
269 AGGCAAGCCAACCGGCTGGTAATCCATCTGCATACCCACAATACAGGTAA 318

 46 lyAspHis...HisThrHisGluHisAspGluSerValSerHisValGly 61
319 TAGAGCACAGCTCCACTAGTAGGCAC.............CACATAGCC 353

 62 LeuGlnAla......HisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
354 CGCTACGCCCTGGTGCATCTACCATCCGCACTTCAG...CACAGAACTGC 400

 76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
401 CCAACCAAGCATCGTACCACACGACCGGGACGCCCACAGGCCCCAATCAC 450

 93 snTyrLeuPro...........ProMetSerGln.........LeuLeu 102
451 CATGGACTCCTCGAACTGCACGCCCAGCATCACAACCTCAGGACACCATG 500

103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
501 ACCGGCGCTACATACAAGAATGCACCAGGACACCGGCCCCACACCACACC 550

119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136
551 CAGCAGACACGTTAAGGCAGCTCCGCACCCCACG...CCACTACAACAAC 597

136 spLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerVal 152
598 CCAGGCGCAACTATGGCCCGTGTCCTAATGGCACTAGGGTGATATCCACG 647

153 Tyr.................ArgAsnProGluLeuAsnGlnCysAlaGl 163
648 TACCCCATCAAGCGCCCTTTCCGTGGTCCTTCACTCCCACGCCGTGGCTG 697

163 yGlyAlaAlaMetSerLysHisLeuThrAsnSer............... 174
698 CGCAATGGCACACCGACCCCATTGGACACATTCACCCCATGGCTCTACCA 747

175 .....AlaIleAspIleTrpValPro 181
748 CTACACCCATGTGATTGTGGCTTCCG 773
```

seq_name: gb_est2:BE426799

seq_documentation_block:

```
LOCUS       BE426799                631 bp    mRNA    linear   EST 24-JUL-2000
DEFINITION  WHE0332_C08_F16ZS Wheat unstressed seedling shoot cDNA library
            Triticum aestivum cDNA clone WHE0332_C08_F16, mRNA sequence.
ACCESSION   BE426799
VERSION     BE426799.1  GI:9424642
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0332_C08_F16"
                     /clone_lib="Wheat unstressed seedling shoot cDNA library"
                     /tissue_type="Etiolated shoot"
                     /dev_stage="Five day old seedling"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
                     , germinated and grown aseptically in the  dark at room
                     temperature on filter paper with water, nystatin and
                     cefotaxime in covered crystallization dishes. Shoots were
                     harvested.  The tissue, total RNA, and poly(A) RNA were
                     prepared, a cDNA library  was made, and the cDNA clones
                     were in vivo excised to give pBluescript  phagemids in the
                     TJ Close lab (Choi, Close, Fenton) at the University  of
                     California, Riverside. Plasmid DNA preparations and DNA
                     sequencing  were performed in the OD Anderson lab (all
                     other authors)."
BASE COUNT      126 a    213 c    163 g    129 t
ORIGIN
```

alignment_scores:

```
          Quality:   92.00            Length:    220
            Ratio:  1.122              Gaps:     11
 Percent Similarity:  37.273   Percent Identity: 22.727
```

alignment_block:

US-09-674-779-2 x BE426799 ..

Align seg 1/1 to: BE426799 from: 1 to: 631

```
 23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
  6 CCCCTGCCTGCCAGCCCCCAAATCTCCCCT.........CCCTCTGTCGC 46

 39 uIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluS 56
 47 TCTC..............................GAATCTTCCCCGA 63

 56 erValSerHisValGlyLeuGln..................AlaHisPhe 66
 64 GCATCGACCATGTCGGCCTACTGCGGAAAGTACAAGGATGAGCTCATCAA 113

 67 GluThrTrpLeuGlnMetHis............HisAlaThrLysGlnGl 79
114 GAACGCTGCCTACATTGGCACCCCTGGCAAGGGTATCCTTGCTGCCGACG 163

 79 uValValArgTyr......GlnAlaTyrLeuGlnSerArg.......... 90
164 AGTCCACCGGCACCATCGGCAAGCGCTTCGCCAGCATCAACGTTGAGAAT 213

 91 .....LeuGlyAsnTyr 94
214 GTTGAGGACAACCGTCGCGCCCTCCGTGAGCTCCTCTTCTGCACCCCTGG 263

 95 LeuProProMetSerGlnLeuLeuThrThrAlaArg.............. 106
264 AGCCCTCCAGTACCTCAGCGGCGTGATCCTCTTCGAGGAGACCCTGTACC 313

107 ..........SerTrpGlnAlaCysGlyHisGluProTyr......GlnL 118
314 AGAGCACCAAGGGTGGCAAGCCCTTCGTCGACATCCTCAAGGAGGGCAAC 363

118 euProProGluHisLeuTrpGlyGln....................... 126
```

```
     364 GTCCTCCCTGGCATCAAGGTGGACAAGGGTACCGTCGAGCTCGCTGGAAC 413

     126 ..................................................... 126

     414 CAACGGTGAGACCACTACCCAGGGCTTTGATGACCTTGGCAAGCGCTGTG 463

     127 .......................IleValProThrLeuHisLeuTyrG 135
                                :::      :::            |
     464 CCAAGTACTATGAGGCTGGTGCCCGCTTTGCCAAGTGGCGTGCTGTCCTC 513

     135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
         |||||       |||:::|||                       :::
     514 AAGATTGGCCTCACCGAGCCATC.......................ACA 539

     152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
         ::::::::::::||||||         |||:::|||:::::::::::::
     540 GCTTTCCATCGACCAGAA.........TGCTCAGGGTCTGGCTCGTTATG 580

     168 rLysHisLeu 171
             ||||||
     581 CCATCATCTG 590
```

seq_name: gb_est1:BB821040

seq_documentation_block:

LOCUS BB821040 440 bp mRNA linear EST 19-NOV-2001

DEFINITION BB821040 RIKEN full-length enriched, mammary gland RCB-0526 Jyg-MC(A) cDNA Mus musculus cDNA clone G830010A14 3', mRNA sequence.

ACCESSION BB821040

VERSION BB821040.1 GI:16993669

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 440)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .440
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G830010A14"
/clone_lib="RIKEN full-length enriched, mammary gland RCB-0526 Jyg-MC(A) cDNA"
/tissue_type="mammary gland"
/cell_line="RCB-0526 Jyg-MC(A)"

BASE COUNT 124 a 132 c 98 g 86 t

ORIGIN

alignment_scores:
Quality: 91.50 Length: 184
Ratio: 1.204 Gaps: 5
Percent Similarity: 41.304 Percent Identity: 20.652

alignment_block:
US-09-674-779-2 x BB821040 ..

Align seg 1/1 to: BB821040 from: 1 to: 440

```
      11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
         |||:::||||||      ::::::   ::::::         ||||||::
      26 ACCATCATATCACCTGCTGTGCTCAAGGCTGAGCTTCAGAGCCCTACTCA 75

      27 nProGlnValSer.........ProIleLysThrProSerVal......L 39
         :|||:::   |||         |||      :::|||::::::
      76 GCCCAGCAGCTCTGCCTTCCCACCGGCTACCAGTCCCACTGCCCCCGAGG 125

      39 euIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspGlu 55
            ::::::      :::   |||      |||::::::::::||||||
     126 CCTACAGCCAGTACCCTGTTCCTGATGTCTCTACTTACCAATATGATGAG 175

      56 SerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMe 72
         :::   ::::::
     176 ACATCTGGCTAC..................................... 187

      72 tHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnS 89
                                                 |||   :::
     188 .....................................TACTATGACC 197

      89 erArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAla 105
         :::   |||   |||       |||   ||||||          :::|||
     198 CCCAGACTGGTCTATACTATGATCCCAATTCCCAGTATTACTACAATGCT 247

     106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHi 122
         :::|||   |||
     248 CAGAGTCACCAG..................................... 259

     122 sLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysS 139
                                  :::||||||   |||   :::
     260 ..........................TACCTATACTGGGATGGGGAGC 281

     139 erArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsn 155
           |||      :::|||||||      |||      :::   :::::::::
     282 GGCGGACCTACATACCTGCCTTGGAGCAGTCTGCTGATGGACATAAGGAC 331

     156 ProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuTh 172
                 ::::::      |||            :::||||||   ||
     332 ACGGGGGCATCATCAAAGGAGGGAAAAGAGAAGAAAGAGAAGCATAAGAC 381

     172 rAsnSerAla.................IleAspIleTrpValProAspL 183
         |:::::::|||                :::|||   |||:::|||:::|
     382 AAAGACAGCCCAACAGATTGCCAAGGACATGGATCGGTGGGCACCCAGTC 431

     183 eu 183
         ||
```

```
     432 TC 433

seq_name: gb_est2:BJ124438

seq_documentation_block:
LOCUS       BJ124438                630 bp    mRNA    linear   EST 23-JAN-2002
DEFINITION  BJ124438 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1311c10 5', mRNA sequence.
ACCESSION   BJ124438
VERSION     BJ124438.1  GI:18284578
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1311c10"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
BASE COUNT      158 a    148 c    139 g    185 t
ORIGIN


alignment_scores:
            Quality:   91.50            Length:    212
              Ratio:  0.915              Gaps:      9
 Percent Similarity:  47.170   Percent Identity: 23.585

alignment_block:
 US-09-674-779-2 x BJ124438   ..

 Align seg 1/1  to: BJ124438  from: 1  to: 630

      31 SerProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAs 47
         |||     :::||||||||||:::         :::|||::::::
      15 TCTTCTACGGAGACACCATCTTTGGAATCAAGACAGGACTCTATG..... 59

      47 pHisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnA 64
          ||||||:::      |||   ::::::::::|||
      60 .CACCACTCTTCGCTCCACCATCACAACATTTCA.............. 92

      64 laHisPheGluThrTrpLeuGlnMetHis.............HisAla 75
                    |||::::::||||||                 :::
      93 ..........CTTTGGTTTTCAATGCATTCGTTATGATGACTGTCTTCA 131

      76 ThrLysGlnGluValValArgTyr..................... 83
         ::::::   :::||||||:::
     132 ATGAGATCAATGCTCGTAAGGTTCACGGAGAGCGTAACGTGTTTAAAGGA 181

      84 .....................................GlnAlaTyrL 87
                                             ::::::   |
     182 TTGGCATCTAACCGAGTTTTCTGTGTCATTTGGGTTACAACATTCATTGC 231

      87 euGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeu..T 103

         ||:::|||   |||   |||         |||   ||||||||||||
     232 TCAAATCATTATTGTTCAATTCGGAGGTGCCTGGTTCTCAACTGCTCCAC 281

     103 hrThrAlaArgSerTrpGlnAlaCys.........GlyHisGluProTyr 116
           |||   :::::::|||   :::|||         |||   :::
     282 TTACGCTTCAACAATGGATTGTATGCCTCGTTCTTGGATTCTCAACTCTT 331

     117 GlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLe 133
                         :::|||||||||||||||   |||:::
     332 ..................ATCTGGGGACAAATTGTGGCAACTATCCCATC 363

     133 uTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleA 150
           :::      |||:::     :::     :::::::::|||
     364 AAAGAAGCTACCAAAGGCCTGGAAAGTTGGAAAGGGAGAAGTTCAACCAG 413

     150 rgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAla 166
           :::::::::   |||   :::   |||      |||   :::   |||
     414 CCAATCTTCACATCAATGGAGACTACAATGTCCGTGCTCGTTCCCGTGCG 463

     167 MetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLe 183
         ::::::      |||   :::|||:::   ::::::||||||      ::
     464 GTTACT......CTTCGCAGATCTGGAAAGTCACTCTGGGTTAGAGGAAT 507

     183 uGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnT 200
         :   |||   ::::::   |||      |||:::             :
     508 GTTCATTATTGGAAACCATCTCCGCGTTCTTCGT............GCTT 545

     200 yrTrpLeuGluHisGlyGluAsnGlnAsnPheGly 211
         ::   :::|||   :::|||:::      ||||||
     546 TCGGAATGGAAAAGAGTGAAAAGGCTGCCTTCGGA 580

seq_name: gb_est1:AL536956

seq_documentation_block:
LOCUS       AL536956                941 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  AL536956 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF023YJ08 5
            prime, mRNA sequence.
ACCESSION   AL536956
VERSION     AL536956.1  GI:12800449
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 941)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .941
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DF023YJ08"
                     /clone_lib="LTI_FL013_FBrn1"
                     /dev_stage="pooled tissue from post conception fetuses (20
                     week, 24 week and 26 week)"
                     /lab_host="DH10B"
                     /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
                     cDNA was primed with a NotI-oligo(dT) primer. Five prime
                     end enriched, double-stranded cDNA was digested with Not I
                     and cloned into the Not I and Eco RV sites of the
                     pCMVSPORT 6 vector.  Library was constructed by Life
                     Technologies.  Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      205 a    223 c    288 g    224 t      1 others
ORIGIN
```

```
alignment_scores:
            Quality:   91.50              Length:    242
              Ratio:  0.897                Gaps:     12
 Percent Similarity:  42.149   Percent Identity: 25.620

alignment_block:
 US-09-674-779-2 x AL536956/rev   ..

 Align seg 1/1  to reverse of: AL536956  from: 1  to: 941

      21 SerAlaProIleProThrAsnProGlnValSerProIleLysThrProSe 37
         |||:::   :::|||:::||||||:::            :::   |||||
     924 AGCAGTCAGTTGCCCTCAAATCCCCGC............AATCTCCCCAG 887

      37 rValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisA 54
         ||||:::|||:::   ::::::::::|||      |||      :::|||:
     886 CGTCATCATATCGTTGCAGCGCTTGGGCACAAAGCACATGACCTCCCACA 837

      54 sp.....................GluSerValSerHisValGlyLeuGln 63
         ::                    :::      |||   |||   |||:::
     836 GCTTGCTCTAGGTCTGCAGTATCCATCCCAGCTCTATTGTAATACTTGAG 787

      64 AlaHisPheGluThrTrpLeuGlnMetHis.........HisAlaThrLy 77
            :::|||      |||   :::|||          |||
     786 AAAATCCTTGAGCAGCAGCTGGTATTTCATGATCCGCTGCACAGGTTTGA 737

      77 sGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnT 94
          :::||||||||||:::   ||||||      |||
     736 TGAGGAGGTCGTTCAGCTGCAGGCGGTGCCCCAG..............C 702

      94 yrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrp.GlnAl 110
           ||||||         :::|||||||   |||:::   :::|||    ||
     701 TGCTGCCG.........GAGCTCCTGGGGACAGGGACAGGGTGGTTATGC 661

     110 aCysGlyHisGluProTyr......GlnLeuProProGluHisLeu.... 123
         |:::         |||         ::::::||||||:::||||||
     660 AGCCTGGAAAGCCCCAGTTCTGTCCTCCTTCCCACCAAGACATCTCAGCT 611

     123 ..................................................... 123

     610 ACTGACCTCAAAGTAGCTGTCCCCAAACTCTGACACCACATGCTCTGACT 561

     124 TrpGlyGlnIleValProThrLeuHisLeuTyr.GlnAspLeuLysSerA 140
         |||:::              |||   |||   ||| |||
     560 TGGGCTTATTCTGACAGTACACCACATACATATGCAGCCGGCGCTCCTAA 511

     140 rgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnPro 156
           |||::::::      :::   |||
     510 CAGGGGTAATGTTTAAGAGAGAGCAG..................... 485

     157 GluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAs 173
                       ||||||   :::         ||||||   ::
     484 .................GGTGGGCAAGGAGGGAGGACGCATCTCCAGAG 453

     173 nSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAla. 189
         :                  :::|||        :::      ||||||
     452 AATCCCAGCCCATTCCTAAGCCCCGCCCCAGGCCCTGCCCCCTCAAGCCT 403

     190 .............LeuTyrGluLeuGlnAsn.ArgLeuCysGlnTyrTr 201
                      :::   |||      ||| |||:::   |||
     402 CACGTGTTTGATGAATAGCTGAGCCAGCCAATCAGGATCTTTCAGACACC 353

     201 pLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrG 218
         |||         ::::::         |||:::|||      |||:::|
     352 GTTGTAGCTCTTGCAAGAAATAGCTGGAGGGGTTGGTAGAGAAGCAGCAG 303

     218 lyAlaIleHisLeuAspThr 224
         ||   ||||||:::   |||
     302 GTCAAATACATTTCACAACC 283

seq_name: gb_est2:BG848755
```

```
seq_documentation_block:
LOCUS       BG848755                673 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  1024022F10.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG848755
VERSION     BG848755.1  GI:14229939
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      147 a    195 c    187 g    144 t
ORIGIN


alignment_scores:
            Quality:   91.00              Length:    116
              Ratio:  1.655                Gaps:      3
 Percent Similarity:  47.414   Percent Identity: 24.138

alignment_block:
 US-09-674-779-2 x BG848755   ..

 Align seg 1/1  to: BG848755  from: 1  to: 673

       7 TyrPheIleThrThrLeuIleSerSerMetLeuValAlaCysSer..... 21
         :::   :::::::::   ::::::      |||   :::|||:::
     348 TTCGTCCTTTCAGCAGCTTTCACCTTAAACTTGCGCAGTTGCGCTCCCGT 397

      22 .AlaProIleProThrAsnProGlnValSerProIleLysThrProSerV 38
          ||||||:::|||   :::|||||||||          ::::::
     398 AGCTCCGGTCCCTCAAGAGCCACAGGTTATCAGCAGTCGATCAGTGTGTG 447

      38 alLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAsp 54
            :::      |||   :::||||||   |||      :::|||
     448 ATGCGGTGGATGACGATGCGGTCGGCGACGGACACATCATCCAACACGCC 497

      55 GluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGl 71
```

```
      :::         |||      |||   :::      :::   |||   ::
    498 CAGATGGAACCGCACAGCCGACTGTGGTCGGCCGC.CGGCGCTGGAGACG 546

     71 nMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuG 88
        ::::   |||   ::::::
    547 CCTGGATCACCCCAGCCGTCCACCACCG..................... 574

     88 lnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThr 104
                  |||   :::|||||||:::                |||
    575 ...........CCGAATGTAGTCCCGCCTCTGCCACTGTGTCGAGGAACA 613

    105 AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro 120
        |||   :::|||   :::|||||      |||:::|||      |||
    614 GCTGCAGGGTGGCTGGGTTGCGGC...GCACCGTGGCAATGTAGGCCA 658
```

seq_name: gb_gss:CNS01Y2Y

seq_documentation_block:

```
LOCUS       CNS01Y2Y                886 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            217L13 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL172483
VERSION     AL172483.1  GI:7810540
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 886)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 886)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .886
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="217L13"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG217CF07LP1~end : T7"
BASE COUNT      218 a    206 c    226 g    235 t      1 others
ORIGIN
```

alignment_scores:

```
           Quality:   91.00            Length:     90
             Ratio:  1.685              Gaps:      5
 Percent Similarity:  60.000   Percent Identity: 32.222
```

alignment_block:

```
 US-09-674-779-2 x CNS01Y2Y/rev   ..

 Align seg 1/1  to reverse of: CNS01Y2Y  from: 1  to: 886
```

```
      6 GlnTyrPheIleThrThrLeuIleSerSerMetLeu.........ValAl 19
        :::|||   :::   :::|||      |||:::|||         |||||
    449 GAATACACCGTCCTCGCCCTCCAGCCCTCTGTATTACAGCACCGGGTTGC 400

     19 aCysSerAlaProIleProThrAsnProGlnValSerProIleLysThrP 36
        |   :::   |||   |||::::::   |||      |||:::   :::|
    399 GCATACATACCCCCGACCATCCGAAGGGCAGGAACCTCCCCTC...GTCC 353

     36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGlu 52
        ||   :::   ::::::|||:::|||      |||||||||   |||
    352 CAGTTCTCTACGTTAACAAAGAAAAGCCCCGCGATCATCATGAACACCCA 303

     53 HisAspGlu......SerValSerHisValGlyLeuGlnAlaHisPheGl 67
        |||   |||      :::|||:::||||||   ::::::||||||
    302 CACCATGAGGTTAATGACGTTCAGCACGTT...GTTGAAGCCCAC..... 261

     67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
                ::::::   |||:::   :::::::::      :::   |
    260 ..........AGAGTTCTTCACTCCCAGAGCCACGATCACCGTCACTAGC 221

     84 lnAlaTyrLeuGlnSerArg 90
        ||   |||   ||||||:::
    220 AGCGATATCAGCAGAGCCAG 201
```

seq_name: gb_est2:BE779469

seq_documentation_block:

```
LOCUS       BE779469                940 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601464207F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3867816 5',
            mRNA sequence.
ACCESSION   BE779469
VERSION     BE779469.1  GI:10200667
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 940)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9614  row: m  column: 01
            High quality sequence stop: 639.
FEATURES             Location/Qualifiers
     source          1. .940
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3867816"
                     /clone_lib="NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      214 a    267 c    276 g    182 t      1 others
ORIGIN
```

alignment_scores:

```
           Quality:   91.00            Length:    250
             Ratio:  0.812              Gaps:     13
 Percent Similarity:  44.800   Percent Identity: 22.400
```

```
alignment_block:
 US-09-674-779-2 x BE779469/rev   ..

 Align seg 1/1  to reverse of: BE779469  from: 1  to: 940

      20 CysSerAlaProIleProThrAsnProGlnValSerProIleLysThrPr 36
         |||:::||||||:::|||   :::   |||:::      :::
     774 TGTGGGGCACCATTGCCCTTTACCTATCAGTTATTTAGGTTGCCCCCCCA 725

      36 oSerValLeuIleThrLysAspLysIleGlyAsp...HisHisThrHisG 52
          :::      ::::::   |||   ||||||           |||||||
     724 AAATCGAGGTTTCTCCGCAGATGTCATAGGTGCCCTCATGTCTACCCATG 675

      52 luHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe..... 66
         ||      ::::::   |||   |||:::|||   :::::::::
     674 AAGGCCCATCAGGGTGCTCCACGGTTGCTCTGGTTGTTGAGATGGATTTT 625

      67 .....GluThrTrpLeuGlnMetHisHisAlaTh 76
                         |||||||||      ::::::::::|||
     624 TAGGGCTCAACTACAGCTGTGGAAACCTGGGCGCCGGTAAATAGAGCAAT 575

      76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
          :::   :::
     574 CCAGCTTGAC..................................... 565

      93 snTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGln 109
           :::||||||                 :::|||   |||      :::
     564 ..TTCTTGCCA.............TAATCAACTGAGAGACGTTCCATG 532

     110 AlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrp..... 124
         :::         :::|||      :::|||||||      ||||||
     531 AGCAGCGAGGTGAACCCAGAACCAGTTCCCCCACCAAAGCTGTGGAAAAC 482

     125 ..........................GlyGlnIleValProThrL 131
                                  ||||||::::::   ::::
     481 CACAGAAGCCCTGGAGACCCGTGCACTGGTCGGCCAGCTTGCGAATTCGG 432

     131 eu.HisLeu...TyrGlnAspLeuLysSerArgGlyIleLeuProAlaAs 146
         :: |||         :::|||||||   ::::::|||:::::::|||:::::
     431 TCCCACACGAGGTCNCATGATCTCCTTGCCAATGGTGTAGTGCCCTCGGG 382

     146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGln.CysAla 162
         :      |||   |||::::::                 ::::::  |||:::
     381 CATAGTTATTGGCAGCATCTTCTTT......GCCTGTGATAAGTTGCTCA 338

     163 GlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTr 179
         |||      :::   ::::::         :::           :::||
     337 GGGTGGAAGAGCTGCGGTAGGTGCCAGTGCGAACTTCATCAATGACTGTG 288

     179 pValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnA 196
         |||||||:::|||:::            |||                 :
     287 GGTTCCAAGTCTACAAACACTGCCCGGGGCA.................C 256

     196 rgLeuCysGln.............................TyrTrp 201
         ::   ||||||                           :::
     255 ATGCTTGCCAGCCCCCGTCTCACTGAAGAAGGTGTTGAAGGAATCATCTC 206

     202 LeuGluHisGly...GluAsnGlnAsnPheGlyLeuGlyLeuTyrAla.. 216
         |||   ||||||      :::   ::::::::::::||||||   |||
     205 CTCCCCCATGGTCTTGTCACTTGGCATCTGGCCATCGGGCTGGATGCCGT 156

     217 .....ThrGlyAlaIleHisLeuAspThr 224
                          :::|||      ||||||||||||
     155 GTTCCAGGCAGTAGAGCTCCCAGCAGGCATTGCCCATCTGGACACC 110

seq_name: gb_est2:BJ111789

seq_documentation_block:
LOCUS       BJ111789               651 bp    mRNA    linear   EST 23-JAN-2002
DEFINITION  BJ111789 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1147d06 5', mRNA sequence.
ACCESSION   BJ111789
VERSION     BJ111789.1  GI:18271832
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk1147d06"
                     /clone_lib="unpublished oligo-capped cDNA library, C.
                     elegans L1 stage"
                     /sex="hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
BASE COUNT      174 a    157 c    127 g    193 t
ORIGIN


alignment_scores:
           Quality:   90.50            Length:    216
             Ratio:  0.887              Gaps:     13
 Percent Similarity:  47.222   Percent Identity: 24.074

alignment_block:
 US-09-674-779-2 x BJ111789   ..

 Align seg 1/1  to: BJ111789  from: 1  to: 651

       2 LysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuVa 18
         :::::::::::::::   |||::::::   |||:::::::::::::   ::
     101 GAACAATGGTCAAGAATATTCTTTGTCATGCTCTTTACCAACTTATCATT 150

      18 lAlaCysSerAlaProIleProThrAsnProGlnValSerProIleLysT 35
         ::::   |||:::                          |||      :::|
     151 ATCTTTGTCATCT.....................TCTTCTACGGAGA 176

      35 hrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHis 51
         ||||||||:::         :::|||::::::      |||||||:::
     177 CACCATCTTTGGAATCAAGACAGGACTCTATG......CACCACTCTTCG 220

      52 GluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluTh 68
            |||   ::::::::::|||
     221 CTCCACCATCACAACATTTCA.........................CT 243

      68 rTrpLeuGlnMetHis..............HisAlaThrLysGlnGluV 80
          |||:::::||||||               :::   ::::::   :
     244 TTGGTTTTCAATGCATTCGTTATGATGACTGTCTTCAATGAGATCAATGC 293

      80 alValArgTyr..................................... 83
         ::|||||||:::
     294 TCGTAAGGTTCACGGAGAGCGTAACGTGTTTAAAGGATTGGCATCTAACC 343

      84 .........................GlnAlaTyrLeuGlnSerArgLe 91
                                  ::::::   |||:::|||   ||
     344 GAGTTTTCTGTGTCATTTGGGTTACAACATTCATTGCTCAAATCATTATT 393

      91 uGlyAsnTyrLeuProProMetSerGlnLeuLeu..ThrThrAlaArgSe 107
         |   |||         |||   ||||||||||||     |||   :::::
```

```
     394 GTTCAATTCGGAGGTGCCTGGTTCTCAACTGCTCCACTTACGCTTCAACA 443

     107 rTrpGlnAlaCys.........GlyHisGluProTyrGlnLeuProProG 121
         :|||   :::|||         |||   :::
     444 ATGGATTGTATGCCTCGTTCTTGGATTCTCAACTCTT............. 480

     121 luHisLeuTrpGlyGlnIleValProThrLeu.His......LeuTyrGl 135
              :::|||||||||||||||   |||::: |||         |||||
     481 .....ATCTGGGGACAAATTGTGGCAACTATCCCATCAAAGAAGCTACCA 525

     135 n.........AspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnI 149
         |           |||:::         :::   |||||||||   :::|
     526 AAGGCCTGGAAAGTTGGAAAGGGGAGAAGTTCAACCAGCCAATCTTCACA 575

     149 leArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla 165
         ||            |||   :::   |||      |||   :::
     576 TC..............AATGGAGACTACAATGTCCGTGCTCGTTCCCGT 610

     166 AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpVal 180
         |||::::::      |||   :::|||:::   ::::::||||||
     611 GCGGTTACT......CTTCGCAGATCTGGAAAGTCACTCTGGGTT 649

seq_name: gb_est2:W80001

seq_documentation_block:
LOCUS       W80001                  656 bp    mRNA    linear   EST 25-JUN-1996
DEFINITION  me88h09.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:402689 5' similar to gb:X70067 M.musculus RNPS1 mRNA
            for RNA/DNA-binding protein (MOUSE);, mRNA sequence.
ACCESSION   W80001
VERSION     W80001.1  GI:1390972
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:246457
            Seq primer: ETPrimer
            High quality sequence stop: 346.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:402689"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT      138 a    193 c    161 g    164 t
ORIGIN


alignment_scores:
            Quality:   90.50            Length:    168
              Ratio:  1.191              Gaps:      9
 Percent Similarity:  45.238   Percent Identity: 28.571

alignment_block:
 US-09-674-779-2 x W80001/rev   ..

 Align seg 1/1  to reverse of: W80001  from: 1  to: 656

      69 TrpLeuGlnMet.........HisHisAlaThrLysGlnGluValValAr 82
         ||||||   ::         ::::::||||||      |||      ||
     604 TGGCTTACAATTCTAGGAAAGTACAGAGCCACCGGGACAGAA......AG 561

      82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetS 99
         |||||||         :::|||      |||      :::   |||:::|
     560 GTACCAATTTGTGGAGGAATCCTTCAAAGGT......TTTAGCCCTCTAA 517

      99 erGlnLeuLeuThrThrAlaArgSerTrpGlnAla.............. 110
         ||   |||   :::   :::|||:::|||:::|||
     516 GT...TTAAGCGTACTCAGCAGGGCATGGTCGGCCCGAACAGAAGAGCAG 470

     111 ..........................CysGlyHisGluProTyrGlnLe 118
                                     :::|||   :::|||      ||
     469 TGTGGTCCTTTGCACAGGTGTGGCACAGCGGGCGCTAGACCAGGCTGTCT 420

     118 uProProGluHisLeuTrpGlyGlnIle.................ValP 129
         |         |||||||||   ::::::                  :::
     419 GGTTCTGGGTCATCTGTGGAAATCCGTGTGCTGCCATAGGTTCCAGCTGG 370

     129 roThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla 145
             |||      ||||||:::||||||:::     :::   |||
     369 CTCTGCTACCAGCCTATCAAGAACTGAAGGGATGAACTTTCTAGCCAGAA 320

     146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
         |||                  ||||||   |||      :::|||
     319 AAC.................TACAGGGCACCTCTATGGGAGCAG..... 293

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
                                        |||||||||:::
     292 .....................CCGTGAACCAACAGTAGCTCTACAAGCC 265

     179 rpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsn 195
             |||::::::|||      ||||||   ||||||   :::::::::
     264 AGAACCCTTCTGTAGAGGAGATATCTCAACTGCTGTACCCAGTAAGGGAT 215

     196 ArgLeuCysGln......TyrTrpLeuGluHisGlyGluAsnGlnAsnPh 210
                 |||         |||||||||         :::|||:::::
     214 CCTTCTCTACAGACCCTCACTTGGCTAGAGAAAAGGACAAAACAGAGTCT 165

     210 eGly 211
         :|||
     164 GGGG 161

seq_name: gb_est2:BG165572

seq_documentation_block:
LOCUS       BG165572                768 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602343723F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4453828 5',
            mRNA sequence.
ACCESSION   BG165572
VERSION     BG165572.1  GI:12672275
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 768)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10244  row: n  column: 05
            High quality sequence stop: 693.
FEATURES             Location/Qualifiers
     source          1. .768
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4453828"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      168 a    186 c    215 g    199 t
ORIGIN


alignment_scores:
            Quality:   90.50             Length:    137
              Ratio:  1.560               Gaps:      8
 Percent Similarity:  42.336   Percent Identity: 28.467

alignment_block:
 US-09-674-779-2 x BG165572/rev   ..

 Align seg 1/1  to reverse of: BG165572  from: 1  to: 768

      21 SerAlaProIlePro.................ThrAsnProGlnValSe 31
         ||||||   ||||||                 :::||||||:::
     438 TCTGCTACAATCCCCCCCAGTCTCCTGATGTCAAGCAATCCTTCGTGACT 389

      31 rProIleLysThrProSerVal.........................L 39
          |||      |||   |||                           :
     388 TCCCTATGCCACAAGGAGCTTTAAAGTCCTATACCACACGCAGGCACACA 339

      39 euIleThrLysAspLysIleGlyAspHis...HisThrHisGluHisAsp 54
         ::   |||         |||      |||   |||||||||   |||
     338 TGTGCACGCACACGCACATACACACACACACACACACACACACACCCT 289

      55 GluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGl 71
         |||||||:::   |||:::                  :::   :::|||::
     288 GAATCTGCTCTCCATGCC.................TCAGATTATCTCAG 257

      71 nMetHisHisAlaThrLysGln........................... 78
         ::::|||||||||||||   |||
     256 TATACACCATGCAACAATGCAGCATTCACAAAACTTTTTTGTTGCTTTTT 207

      79 ..GluValValArgTyrGlnAlaTyrLeuGlnSerArgLeu......... 91
           ||||||:::|||   ::::::      |||:::   |||
     206 AGGAAGTAGCGAGGAAGGAAAGCAAGCAGCAGGATCCCCTAGAGAGTTTA 157

      92 .......................GlyAsnTyrLeuProProMetSerGl 100
                                |||::::::|||      ::::::||
     156 GTCTTTGGTTTCTAAGTTTAAAGGGGGGATTGGCTTCAGAGCTTGGAGCA 107

     100 n......LeuLeuThrThrAlaArgSerTrpGln............AlaC 111
         |       :::          |||||||||             :::
     106 AGACAGAAGAGTCGACGGACGGATGAGCTGGCAAGGGAGAAGGGAGTCTC 57

     111 ysGlyHisGlu 114
           |||||||||
      56 TGGGGCATGAG 46

seq_name: gb_gss:BH269012

seq_documentation_block:
LOCUS       BH269012                807 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  CH230-66O5.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-66O5, DNA sequence.
ACCESSION   BH269012
VERSION     BH269012.1  GI:17181318
KEYWORDS    GSS.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 807)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
            ,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
            Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-66O5.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 66  row: O  column: 5
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .807
                     /organism="Rattus norvegicus"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-66O5"
                     /clone_lib="CHORI-230 Segment 1"
                     /sex="Female"
                     /cell_type="Brain"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
BASE COUNT      211 a    197 c    197 g    202 t
ORIGIN


alignment_scores:
            Quality:   90.50             Length:    146
              Ratio:  1.293               Gaps:      8
 Percent Similarity:  47.945   Percent Identity: 24.658

alignment_block:
 US-09-674-779-2 x BH269012/rev   ..

 Align seg 1/1  to reverse of: BH269012  from: 1  to: 807

      92 GlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTr 108
```

```
     |||         |||   :::||||||:::|||   |||:::|||:::||
 465 GGGGCTAATGCGCCTTCTCTCTCCCAGATTTTG...ACAGTGAGGAACTG 419

 108 pGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuTrp. 124
     |:::          :::   ::::::   ||||||::::::||||||
 418 GAGAAGGAGTTGGACATCCTCCTTCAGGACACCACCAAAGAACCTTTGGA 369

 125 ...................GlyGlnIleValProThr............. 130
                          :::::::::||||||
 368 CCTGCTCGAAAACCCCCATGAGACGCTTTATACCAACAGTGTGCCTAAGC 319

 131 ..............LeuHisLeuTyrGlnAspLeuLysSerArgGlyIl 142
                  |||:::|||      :::|||::::::
 318 CTAGGATCTTGGATGCTGAACTTGAAGCTGAACTTGAGAAAC........ 277

 142 eLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuA 159
         |||   :::   ::::::||||||   :::      |||
 276 .TATCCTTATCAGAAGGAGGTACGGAGCTATTTCCCATGTCCT....... 235

 159 snGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThr......... 172
                               :::|||:::|||
 234 ............................AGACACGTGACTCTGGCTGAT 214

 173 ......AsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSe 187
           :::::::|||   :::      :::|||:::      |||:::||
 213 AGGAGGAGCCAGGCATAGAACTGACAGATTCCATCAAAAACAATCAGATC 164

 187 rGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeu.Glu 203
     |                |||||||          :::   ||||||
 163 C.................CAAAACCCTGCAAGGGAGAAGTGGTTATGTA 132

 204 HisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyr 215
     |||||||||   ::::::   :::||||||   :::
 131 CATGGAGAAGCAGATGGTCAAGCTTTGGGATGCTTT 96

seq_name: gb_gss:AQ744833

seq_documentation_block:
LOCUS       AQ744833                 815 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_5506_A2_G06_SP6 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1082 Col=12 Row=M, DNA sequence.
ACCESSION   AQ744833
VERSION     AQ744833.1  GI:5522355
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 815)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1082  row: M  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 815.

FEATURES             Location/Qualifiers
     source          1. .815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1082 Col=12 Row=M"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      222 a     77 c    301 g    214 t      1 others
ORIGIN


alignment_scores:
           Quality:   90.50              Length:    205
             Ratio:  0.862                Gaps:     10
 Percent Similarity:  51.220   Percent Identity: 22.439

alignment_block:
 US-09-674-779-2 x AQ744833/rev   ..

 Align seg 1/1  to reverse of: AQ744833  from: 1  to: 815

       3 AsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeu...Va 18
         :::::::::::::|||::::::      |||||||||:::   |||   ::
     710 TCCTACTCCTCTTATCTTCTCCCTCTCCTAATATCCACCCATCTCCCACT 661

      18 lAlaCysSerAlaProIleProThrAsnProGlnValSerProIleLysT 35
         ::::   |||   |||:::|||   |||   :::|||:::|||:::
     660 CTCTCACTCCCACCCCCTACCACACAACTCACATGTCACACCCTTCCTAC 611

      35 hrProSerValLeuIleThrLysAspLysIleGly......AspHisHis 49
           |||:::   |||          :::   ::::::      :::|||:::
     610 CCCCAACCTACCTCAACCACCCCTCTCCTGTCTCTCTCCCCTCCCATTAT 561

      50 ThrHisGluHisAspGlu...SerValSerHisValGlyLeuGlnAlaHi 65
             :::|||          :::   :::|||:::   |||   :::
     560 CCCCTCTCACACACCCCCCTTACTTTCACACACCTCACCCTCCTGTCCTA 511

      65 sPheGluThrTrpLeuGlnMet.....................HisHisA 75
         :::|||   |||::::::                        |||:::|
     510 NACCTCAACATCTCTCTCAATCCCCCCTCCTTAATACCTTCCACACCGCG 461

      75 laThrLysGlnGluValValArgTyrGlnAlaTyr.............. 86
         |||||      :::::::::      :::   |||
     460 CTACCCTACCCTCTCTCCTCCCACAATCTCCTTACCCTTACCTCATCTCT 411

      87 .....................LeuGln...SerArgLeuGlyAsnTyrLe 95
                              ||||||   ||||||   ::::::::::
     410 CTACACAACCTCCGAACTTCGCTCCAACACTCCCGACCATCTTCACACTT 361

      95 uProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysG 112
         :||||||         |||              |||:::   :::   :
     360 TCCTCCACGCTTCTACCTC............ACGTCCTTTCTTGGCTTCT 323

     112 lyHisGluProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleVal 128
         ::   :::|||         ||||||:::      |||   :::|||
     322 CTTTATCGCCACCCGCTCCACCCCCATCTTCACACTGGCGTCATATAAGT 273

     129 ProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAl 145
         |||                                :::   ||||||
     272 CCT............................CAATACATTCTCATACT 253

     145 aAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysA 162
          :::   :::|||           ::::::|||:::::::::   |||
     252 GGGTGATAAAATA............AACAGACCCCAGGTTCAT...TGTC 218

     162 laGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIle 178
           :::::::::   ::::::      :::   ::::::   :::   |||
     217 AGAGCAATTCCCGACTTGCTCTGACAATTGGGGAGGAATTGGTGAGAATA 168
```

```
    179 TrpValProAspLeu 183
        |||   |||   |||
    167 TGGAATCCAAAATTG 153

seq_name: gb_est2:BG292887

seq_documentation_block:
LOCUS       BG292887               1055 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602389522F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4501396 5',
            mRNA sequence.
ACCESSION   BG292887
VERSION     BG292887.1  GI:13052166
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1055)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10368  row: 1  column: 05
            High quality sequence stop: 294.
FEATURES             Location/Qualifiers
     source          1. .1055
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4501396"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      361 a    305 c    267 g    122 t
ORIGIN


alignment_scores:
            Quality:   90.50            Length:    209
              Ratio:  0.963              Gaps:      9
 Percent Similarity:  44.976   Percent Identity: 24.402

alignment_block:
 US-09-674-779-2 x BG292887   ..

 Align seg 1/1  to: BG292887  from: 1  to: 1055

      17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProIl 33
         |||   ||||||   ||||||   ||||||:::   ||||||   |||
      84 TTACCCGCCTGCCCAGCCCCTCCTCCCACAGGAACACAGGTACCACCTGC 133

      33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHis.... 48
             |||   :::   :::::|||      :::   |||   :::
     134 ACCCACGAGAGCTTCTTTTCTCACACATGGGGAGACTGGAGCCGAGCCCC 183

      49 ............................HisThrHisGluHisAspGlu 55
                                     ::::::|||||||||
     184 TTGTGTCTTTTGAAATAGGAAGAAAAATGAGTGTTCAGGAACATGAC... 230

      56 SerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMe 72
                |||   :::||||||   |||:::|||::::::      |||::
     231 ......TCCTGGACTGGGCTCTGGGCCCAGTTCAGTGTCTCATCGCAAAT 274

      72 t.HisHis..................................... 74
         : :::|||
     275 CTAGGCACTTTTCGCTTCTATAGTCTATTTAATTGTTTTATTATTTAAGG 324

      75 ......AlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSe 89
                |||||||:::      ::::::   :::         |||||
     325 AAAAACAACACAAAACGGGGGGGGGCGAACCAGAAAATTTCATCCCAGTC 374

      89 rArg.........LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuT 103
         ||||           |||   |||   |||                  |
     375 GAGGGGGCCCCAAACACATAACCGGCTATCCCCGCGATTCTTTTGGGCCA 424

     103 hrThr......AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
         |||||        ::::::|||:::         :::   |||   :::
     425 CCACATGGGGCCCCAAAAAGTGGAGG.........AACGGCCCACTAAAC 465

     118 LeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134
         |||||||:::         |||:::::::::   :::   |||
     466 CAACCGCCTCAGACCCCCACCGGGGACGTGACGCGTAACAACCACGCGTC 515

     134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
          :::   |||::::::   |||           :::   :::
     516 GGAAACGCTGGAAACCACAGGACACACGAGGGACACCCCGGAGCGAGACA 565

     151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGly............ 163
                :::||||||              ||||||
     566 GGAAACCCAAAAACCCA..............GCGGGATGTCACAACATT 600

     164 GlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpVa 180
         ::::::   :::   |||   ||||||::::::
     601 AAGGGGAGCCACACAACCCACACGACAAATAAGACGACCCCAACGGCGCC 650

     180 lProAspLeuGluIleLysSerGln 188
          |||:::   ::::::|||:::|||
     651 ACCATCCACCAACCTCAAAGGACAA 675

seq_name: gb_gss:AZ083675

seq_documentation_block:
LOCUS       AZ083675               594 bp    DNA     linear   GSS 08-MAY-2000
DEFINITION  RPCI-23-22M17.TJ RPCI-23 Mus musculus genomic clone RPCI-23-22M17,
            DNA sequence.
ACCESSION   AZ083675
VERSION     AZ083675.1  GI:7725408
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-22M17.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 22  row: M  column: 17
```

```
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-22M17"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       162 a    139 c    132 g    161 t
ORIGIN


alignment_scores:
            Quality:   90.00              Length:    126
              Ratio:  1.324                Gaps:      7
 Percent Similarity:  53.968    Percent Identity: 28.571

alignment_block:
 US-09-674-779-2 x AZ083675/rev   ..

 Align seg 1/1  to reverse of: AZ083675  from: 1  to: 594

       8 PheIleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIl 24
         |||:::          ||||||||||||||||||    :::
     494 TTTTTACATGGTGAGATTTCCTCAATGTTAGTTTTTGCTTAAATCTCACA 445

      24 eProThr.AsnProGlnValSerProIle.............LysThr 35
          |||||| :::||||||::::::   :::             ::::::
     444 ACCGACACCGCCCCCAATTGCAATCCCTTATATATCTTCTCAACCGTGCC 395

      36 ProSerValLeuIleThr...LysAspLysIleGlyAspHisHisThrHi 51
         ||||||:::     |||   ::::::::::||||||:::::|||   ||
     394 CCATCAGCTCAGTCCACATAACAGCAGTCCATTGGCCAGAATCAT...CA 348

      51 sGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu. 67
         |:::          :::   |||   |||      |||:::|||:::
     347 CTCATCATATGGTCCAATCTTGCATCATGGTGCACCTGCTCAGTTCTCAC 298

      68 .....ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArg 82
               ||||||::::::::         |||:::     :::
     297 AATGAATGTGGCTTAATTTTAGGTTTATGAGGAAGTCAGGT...GCAAGT 251

      83 TyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSe 99
         ::::::::::::::|||||||||
     250 CATAAGACTTGGATGCAGTCCCGG..........................  227

      99 rGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProT 116
                    |||   ||||||::::::|||         |||:
     226 ...............GCACCATCTTGGGACTGCTGCTGTAAGACCCCCT 193

     116 yrGlnLeuProProGluHisLeuTrp 124
         :::::   ||||||         |||
     192 TTCGGGGACCCCCACTCTAGTATTGG 167

seq_name: gb_est1:BB665280

seq_documentation_block:
LOCUS       BB665280                700 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB665280 RIKEN full-length enriched, 0 day neonate eyeball Mus
            musculus cDNA clone E130116F03 5', mRNA sequence.
ACCESSION   BB665280
VERSION     BB665280.1  GI:15403126
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K.,. Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="E130116F03"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     eyeball"
                     /tissue_type="eyeball"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
```

```
BASE COUNT      147 a    223 c    195 g    135 t
ORIGIN


alignment_scores:
            Quality:   90.00             Length:    106
              Ratio:  1.607               Gaps:      5
 Percent Similarity:  52.830   Percent Identity: 29.245

alignment_block:
 US-09-674-779-2 x BB665280/rev   ..

 Align seg 1/1  to reverse of: BB665280  from: 1  to: 700

      78 GlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTy 94
         :::|||:::   |||   :::|||   :::   |||   :::|||:::
     237 AGGGAAGCCCAGAGGATGAAGGCA...GTGACAAGCATCGTGGGAGGGGC 191

      94 rLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaC 111
               |||   |||:::|||                  |||   |
     190 CTGGGGCCCAGGAAGCAGGCTC.....................CAGTTCT 162

     111 ysGlyHis......GluProTyrGlnLeuProProGluHisLeuTrpGly 125
         ||||||||        |||::::::::::||||||:::::|||
     161 GTGGACACCTGCGCGCCCCCTTCCGCGTTCCTCCTTCCTACTTG..... 118

     126 GlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIl 142

     118 ..................................................... 118

     142 eLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuA 159
         |||||||||::::::|||   |||   ::::::   :::|||   |||:
     117 .CTGCCAGCCACGGCCCAGCGGCGAAGGACCTGGGATGGCCCAGGCCTCT 69

     159 snGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAla 175
         ::   ::::::||||||||||||::::::          |||::::::
      68 CA...GCAGGTGGCGGTGCAGCCCTTGGGACGGTTGGCACTTCTGGACCT 22

     176 IleAspIleTrpValPro 181
         |||   |||:::|||
      21 TCCGACAGATGGCTACCT 4

seq_name: gb_est2:BG368308

seq_documentation_block:
LOCUS       BG368308                746 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEi0017J13f Hordeum vulgare 20 DAP spike EST library HVcDNA0010
            (20 DAP) Hordeum vulgare cDNA clone HVSMEi0017J13f, mRNA sequence.
ACCESSION   BG368308
VERSION     BG368308.2  GI:16325190
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 746)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Close,S.J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex 20 DAP spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Mar 8, 2001 this sequence version replaced gi:13257409.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 438
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 562.
FEATURES             Location/Qualifiers
     source          1. .746
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEi0017J13f"
                     /clone_lib="Hordeum vulgare 20 DAP spike EST library
                     HVcDNA0010 (20 DAP)"
                     /tissue_type="20 DAP spike"
                     /lab_host="SOLR"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were grown in the greenhouse at the University of
                     California, Riverside (Fenton, SJ Close, TJ Close). Whole
                     spikes with awns trimmed were collected at 20 DAP (Fenton
                     ). Total RNA was prepared, poly(A) RNA was purified, one
                     primary unamplified cDNA library was made, and 1 million
                     pfu were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids in the TJ Close lab at the University of
                     California, Riverside (Choi). Phagemids were plated and
                     picked at the Clemson University Genomics Institute (CUGI)
                     (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
                     performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
                     , Rambo, Main). The sequence has been trimmed to remove
                     vector sequence and contains a minimum of 100 bases of
                     phred value 20 or above. For more details on library
                     preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      162 a    251 c    189 g    143 t      1 others
ORIGIN


alignment_scores:
            Quality:   90.00             Length:    214
              Ratio:  1.098               Gaps:     11
 Percent Similarity:  38.318   Percent Identity: 22.430

alignment_block:
 US-09-674-779-2 x BG368308   ..

 Align seg 1/1  to: BG368308  from: 1  to: 746

      56 SerValSerHisValGlyLeuGln.................AlaHisPh 66
         |||::::::||||||||||||                   ||||||
      49 AGCATCGACCATGTCGGCCTACTGCGGAAAGTACAAGGATGAGCTCATCA 98

      66 eGluThrTrpLeuGlnMetHis.......................... 73
          |||      |||:::   |||
      99 AGAACGCTGCCTACATTGGCACCCCTGGAAAGGGTATCCTTGCTGCTGAC 148

      74 .....HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGln 88
            |||         ::::::::::::|||   :::          :::
     149 GAGTCCACCGGCACCATCGGCAAGCGCTTCGCCAGCATCAATGTTGAGAA 198

      89 SerArg.................................LeuGlyAsnTy 94
         |||                                  |||      ::
     199 CGTCGAGGACAACCGCCGTGCCCTCCGTGAGCTCCTCTTCTGCACCCCTG 248

      94 rLeuProProMetSerGlnLeu.....................LeuT 103
         :   ||||||:::   |||                        |||
     249 GTGCCCTCCAGTACCTCAGCGGTGTGATCCTCTTCGAGGAGACCCTCTAC 298

     103 hrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyr......Gln 117
               ::::::||||||||||      ::::::|||          |||
     299 CAGAGCACCAAGGGTGGCAAGCCCTTCGTCGACATCCTCAAGGAGGGCAA 348

     118 LeuProProGluHisLeuTrpGlyGlnIleVal................ 128
```

```
          ||||||:::|||      ||||||
     349 CGTCCTCCCCGGCATCAAGGTGGACAAGGGTACCATCGAGCTCGCTGGAA 398

     129 ..............................ProThrLeuHisLeuT 134
                                        |||         |||
     399 CCAATGGTGAGACCACCACCCAGGGCTTTGATGACCTTGGCAAGCGCTGC 448

     134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
           |||   |||      ||||||      |||           :::|||
     449 GCCAAGTACTA......CGAGGCTGGTGCCCG...........CTTCGC 480

     151 SerValTyrArgAsnProGlu.......................... 157
         :::|||      :::|||:::
     481 CAAGTGGCGTGCGGTCCTCAAGATCGGTGCCACCGAGCCATCGCAACTCT 530

     158 ....LeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrA 173
               :::|||:::|||:::::::::::::      ||||||
     531 CCATCGACCAGAATGCTCAGGGTCTGGCTCGCTATGCCATCATCTGCCAN 580

     173 snSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAla 189
           :::      :::|||   |||:::|||:::::::||||||
     581 GAAGAACG......GTCTGGTGCCCATCGTTGAGCCTGAAATCC...... 618

     190 LeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGlu 203
                                 :::|||::::::
     619 ............................TGGTGGATGGAC 630

seq_name: gb_gss:BH070893

seq_documentation_block:
LOCUS       BH070893                533 bp    DNA     linear   GSS 18-JUL-2001
DEFINITION  RPCI-24-275C20.TV RPCI-24 Mus musculus genomic clone RPCI-24-275C20
            , DNA sequence.
ACCESSION   BH070893
VERSION     BH070893.1  GI:14890490
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 275  row: C  column: 20
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .533
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-275C20"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      146 a    144 c    127 g    116 t
ORIGIN


alignment_scores:
            Quality:   89.00             Length:     87
              Ratio:  1.894               Gaps:      3
 Percent Similarity:  54.023   Percent Identity: 32.184

alignment_block:
 US-09-674-779-2 x BH070893   ..

 Align seg 1/1  to: BH070893  from: 1  to: 533

      23 ProIleProThrAsnPro......GlnValSerProIleLysThrProSe 37
         ||||||||||||:::|||      |||:::   |||:::      |||::
      64 CCTATTCCAACCAGGCCACACCTCCAGATAGTGCCACTCCCTGGTCCGAA 113

      37 rValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisA 54
         ::::         :::               :::|||:::      |||
     114 AATATACAAACCATCACACGCATATACACACAGACATGCATGCCCACATA 163

      54 spGluSerValSerHisValGlyLeuGlnAlaHisPheGluThr..... 68
           |||:::   ||||||||||||:::   :::   |||:::|||
     164 CAGAGGAACATAGTCACGTGGGGATCCTCAGTTTCTTCTCTACTGTCTCC 213

      69 ............Trp.LeuGlnMetHisHisAlaThrLysGlnGluValV 81
                      ||| |||:::       :::||||||   ::::::|
     214 TTTCTCTCCTTATGGTCTTTCCTGCTGCTCATCCACCAAGCCCAAAGCAG 263

      81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
         |||||                |||   |||   :::::::::|||
     264 TACGTGGCCCCCAGCAGTGTCAGCACCGTGCCTCTGGCTTTCTAGAACTG 313

      98 MetSerGln 100
         ||||||
     314 TGAAGTCAG 322

seq_name: gb_est1:BB692619

seq_documentation_block:
LOCUS       BB692619                550 bp    mRNA    linear   EST 10-OCT-2001
DEFINITION  BB692619 RIKEN full-length enriched, 2 days neonate sympathetic
            ganglion Mus musculus cDNA clone 7120437K23 3', mRNA sequence.
ACCESSION   BB692619
VERSION     BB692619.1  GI:16019352
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
```

```
          URL:http://genome.gsc.riken.go.jp/
          Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
          ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
           Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new
          genes. Genome Res. . 10 (10), 1617-1630 (2000)
           wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
          Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
          ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
          Hayashizaki,Y.
           RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
          10 (11), 1757-1771 (2000)
           Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
          ,Y. and Hayashizaki,Y.
           Computer-based methods for the mouse full-length cDNA
          encyclopedia: real-time sequence clustering for construction of a
          nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
           Please visit our web site (http://genome.gsc.riken.go.jp) for
          further details.
           e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="7120437K23"
                     /clone_lib="RIKEN full-length enriched, 2 days neonate
                     sympathetic ganglion"
                     /sex="mixed"
                     /tissue_type="sympathetic ganglion"
                     /dev_stage="2 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I.
                     Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
                     ~RNA was provided by Akira Nakagawara, Div. of
                     Biochemistry, Chiba Cancer Center Research Institute,
                     666-2 Nitona, Chuoh-ku, Chiba, 260-8717 Japan, whose
                     assistance we gratefully acknowledge."
BASE COUNT      129 a    115 c    128 g    178 t
ORIGIN


alignment_scores:
            Quality:   89.00             Length:    139
              Ratio:  1.413               Gaps:      7
 Percent Similarity:  45.324   Percent Identity: 23.741

alignment_block:
 US-09-674-779-2 x BB692619/rev   ..

 Align seg 1/1  to reverse of: BB692619  from: 1  to: 550

      15 SerMetLeuValAla...CysSerAlaProIlePro..............  25
         |||:::|||   |||   |||   :::   ||||||
     485 TCTCTACTGAAAGCACTTTGTTTGGTTTTAATTCCACCTCCAGAGTCAAA 436

      26 .......ThrAsnProGlnValSerProIleLysThrProSerValLeuI 40
                 :::|||:::   |||               |||:::|||

      435 GGCAAGACCAAGACCTGACAGCTCTGAGGGCCTTCCCAACAGTGCCCTGC 386

       40 leThrLysAspLysIleGlyAspHis.........HisThrHisGluHis 53
          ::::::   :::|||:::   |||         |||||||||   |||
      385 CCTCAAGGCTTAGAATTAGCACGCACACACACACACACACACACACACAC 336

       54 AspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLe 70
          ::::::   :::|||:::   :::|||   |||   :::|||:::
      335 ACACGCACACACACACACACACACATACAGAGGCACACACACACATACAC 286

       70 uGlnMetHis...HisAlaThrLys........................... 77
          :::::::|||   |||       |||
      285 GAACTTGCACAGGCATGATTTAAAAGGTGCTTGATGTCAGGTTAAAAGAG 236

       78 .........................GlnGluValValArgTyrGlnAla 85
                                   :::|||::::::   |||:::
      235 AAGTTGGGGCCAGTATTAGAGACATTTTGGATGTTATACAGACACAAACT 186

       86 TyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLe 102
          :::   :::|||:::   ||||||:::
      185 CACAAGGAATCAAAAAACGGCAATTTC....................... 159

      102 uThrThrAlaArgSerTrpGln.......................AlaC 111
                  :::   ||||||                        :::|
      158 ....TGGGAAAAATACTGGCAGGAACGGAACTTTTTTTTTTTAACTTCAT 113

      111 ysGlyHisGluProTyr 116
          ||||||||:::   :::
      112 GTGGACATTCTACGTTC 96

seq_name: gb_est2:BG207133

seq_documentation_block:
LOCUS        BG207133               647 bp    mRNA    linear   EST 21-APR-2001
DEFINITION   RST26599 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION    BG207133
VERSION      BG207133.1  GI:13728820
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 647)
  AUTHORS    Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
             Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
             , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
             ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
             ,J., Danzig,J. and Ducar,M.
  TITLE      Creation of genome-wide protein expression libraries using random
             activation of gene expression
  JOURNAL    Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE    21227151
COMMENT      Contact: Scott J. Cain
             Athersys, Inc.
             3201 Carnegie Ave, Cleveland, OH 44115, USA
             Tel: 216 431 9900
             Fax: 216 361 9596
             Email: scain@athersys.com
             High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      128 a    129 c    167 g    222 t      1 others
ORIGIN
```

```
alignment_scores:
            Quality:   89.00              Length:    185
              Ratio:  1.035                Gaps:     12
 Percent Similarity:  46.486   Percent Identity: 27.027

alignment_block:
 US-09-674-779-2 x BG207133/rev   ..

 Align seg 1/1  to reverse of: BG207133  from: 1  to: 647

      25 ProThrAsnProGlnValSerProIleLysThrProSerVal......Le 39
         |||   :::|||||||               |||||||::         ::
     632 CCAGAGAAGCCCCAG..............ACCCCTGAAAGTAGGTGGAT 598

      39 uIleThrLysAspLysIleGlyAspHisHisThrHis...GluHisAspG 55
         :   |||      :::||||||:::   ::::::         |||:::|
     597 TGCCACTGCAGGCTCCATTGGACAGCCAAGGAACTTCTAGCTCCACAGTG 548

      55 luSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGln 71
         |||||:::|||:::       :::   :::       |||   |||||||||
     547 AGAGCCTATCAAACTAGTAAGTCACGAGTGACCATGAGCCCTGGCTTCAG 498

      72 MetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGl 88
         :::         ::::::|||:::   :::       :::       |||
     497 CTC......CCAGTGGAACAGCAGCCAACCAGCCTGGAATACATACTCCT 454

      88 nSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrA 105
          |||:::             ||||||          ||||||
     453 TTCCAAGAGCCAGCCTGCACTACCAGTCCCATGCCAGCTATACCTACTG. 405

     105 laArgSerTrpGlnAlaCys............................. 111
             |||       |||
     404 .....TACTGGACACCCTGCTTCTAGCTCCACAGTGAGAGCCTATCAAAC 360

     112 ..............GlyHisGluProTyr...GlnLeuProProGluHi 122
                       |||||||||:::   |||||||||   |||::
     359 TAGTAAGTCACGAGTGACCATGAGCCCTGGCTTCAGCTCCCAGTGGAGCA 310

     122 sLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu...L 138
         :                  ||||||          |||       |||   |
     309 G..............CAGCCAACCAGCCTGGAATACATACTCCTTTCCA 275

     138 ysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyr... 153
         |||||:::      ||||||:::      |||:::      :::|||
     274 AGAGCCAGCCTGCACTACCAGTCCCATGCCAGCTATACCTACTGTACTGG 225

     154 ..............ArgAsnProGluLeuAsnGln.........CysAl 162
                      |||||||||:::::::|||:::         |||::
     224 ACACCCTGCTGTGAAAGAAATCCAAGAGTGAACAAGACCAAGTCTTGCTC 175

     162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
         :   |||::::::::::         |||   |||:::   :::       :
     174 TTATGGAACTTCTATT.........CTAGGGAATGGGGAAGTTGGTTATT 134

     179 rpVal 180
         :::::
     133 ATATA 129

seq_name: gb_gss:CNS02K1D

seq_documentation_block:
LOCUS       CNS02K1D                1023 bp    DNA     linear   GSS 14-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            144J11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL200938
VERSION     AL200938.1  GI:7859283
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1023)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1023)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1023)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1023
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="144J11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG144CE06SP1-end :
                     PUC-Ori"
BASE COUNT      232 a    273 c    271 g    242 t      5 others
ORIGIN


alignment_scores:
            Quality:   89.00              Length:    306
              Ratio:  0.761                Gaps:     19
 Percent Similarity:  38.235   Percent Identity: 20.915

alignment_block:
 US-09-674-779-2 x CNS02K1D   ..

 Align seg 1/1  to: CNS02K1D  from: 1  to: 1023

      23 ProIleProThrAsnPro.........GlnValSerProIleLysThrPr 36
         |||   |||:::||||||         |||       ||||||
     225 CCCCCGCCATCAAACCCTGTCTCTAAGCAGGATTGGCCCATT........ 266

      36 oSerValLeuIleThrLysAspLysIleGlyAspHis.............. 48
                                  ||||||||||||
     267 .........................ATTGGAGACCACCCTGTAGGAGGAT 291

      49 .....HisThrHisGluHisAspGluSer...................... 56
               |||||||||   |||   ::::::
     292 GCACACACACACACACACACACACACACACGCTTTGTGATAGTGTTTGTT 341

      56 ....................................................... 56

     342 ACTGTATACGGCGGCAGGGCCGTTACGCGTTGAGTYAGCCTCCAGCCCTG 391

      57 .............ValSerHisValGlyLeuGlnAlaHisPheGlu.... 67
                      |||:::         |||   ::::::|||:::
     392 CGTAAAAGCTCCTGTTGCAGTATCGACTCTACCGGGTTATTTTCAGCCAC 441

      68 ..............................................Thr 68
                                                      |||
     442 CGATAATGAAACACCTGGATCAGCTAATAGGTTGGGGGGTCGGAGCTACT 491

      69 TrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAl 85
         |||::::::   |||:::::::::                    ::::::::
```

```
    492 TGGATGGAGTCCCACAACAGCAGC.................TTTAGAGT 523

     85 aTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuL 102
        :   ||||||::::::   |||:::   |||      |||      ::::
    524 WAAACTGCAAAAAAAAAAAAGGTAGAGAACTTAAAAAGATGTATGTTTTCA 573

    102 euThrThrAlaArgSerTrp.................GlnAlaCysGly 112
        ::|||   :::|||   |||                 |||:::   |||
    574 TCACTATGGTGAGA...TGGGTCCACCCCCGGCTGGTTCAGGGGTTTGGA 620

    113 HisGluProTyrGlnLeuPro......ProGluHisLeuTrp........ 124
           :::|||   :::||||||      |||      :::|||
    621 ...AGACCCTCTCGATTACCTGTCGTACCTGTAAAGATCTGGAAAAACAA 667

    125 ................................Gly.GlnIleValPro 129
                                       ||| |||      |||
    668 AAAAAACCACCTCAATTTGCTATCTCACCAATGTGGTACAGTCCGATCCA 717

    130 Thr......LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuPr 144
        :::      :::|||::::::   |||                 :::||
    718 GTCTGTGGCATCCACTTCTTCTTTGAT.................GTCCC 749

    144 oAlaAsnThrGlnIle...ArgSerValTyrArgAsnProGluLeuAsnG 160
        |         ::::::      :::|||   |||      :::   ::::
    750 ACGATATGACGAGGTTCTGGCCACGGTTCAGCGTGTACTCCAGCGTGGAC 799

    160 lnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIle 176
        ::|||      |||   |||::::::  |||   ::::::   |||:::
    800 GCTGTCAGTGAGGACCGGCTCTCGGAGGTCACCAGGTCCGTGTCGCTGTT 849

    177 AspIleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLe 193
        :::          |||
    850 AGCCCGGGTGAGTCCGAT............................... 867

    193 uGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnP 210
                                       |||      :::|||
    868 ................................GGGACTKCTGACAAC. 882

    210 heGlyLeuGlyLeuTyrAlaThrGlyAlaIle..................... 220
          |||   |||:::::::::::::||||||
    883 ..GGAGACGGTATCCACGGCAGCGGCGCCGCCCCTCTCCGTCAGGGTCCG 930

    221 ...HisLeuAspThrGlnGly.PheArgLysTrpGlyAlaGlnPheSerG 236
           |||   ::::::||| ||||||:::||||||      |||:::
    931 CAGTTTCTCCGGGCTGAGGGTGTTCCGGAGTTGGGGGCTGCCCTTCGCGG 980

    236 luThrAsnSerIle 240
           :::|||:::
    981 GGCTGACAAGTGTT 994

seq_name: gb_est2:BG291582

seq_documentation_block:
LOCUS       BG291582               1032 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602385677F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4514778 5',
            mRNA sequence.
ACCESSION   BG291582
VERSION     BG291582.1  GI:13049829
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10403  row: i  column: 19
            High quality sequence stop: 155.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4514778"
                     /clone_lib="NIH_MGC_93"
                     /tissue_type="transitional cell papilloma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      296 a    423 c    114 g    199 t
ORIGIN


alignment_scores:
            Quality:   89.00              Length:    170
              Ratio:  1.219                Gaps:      9
 Percent Similarity:  42.941   Percent Identity: 22.353

alignment_block:
 US-09-674-779-2 x BG291582   ..

 Align seg 1/1  to: BG291582  from: 1  to: 1032

      25 ProThrAsnProGlnValSerProIle..................... 33
         |||   |||   |||:::   |||:::
     166 CCACCTAATCAGCAACTGGTGCCAGTCTAATGGGTTCGGTGCACATAACA 215

      34 .........LysThrProSerValLeuIleThr............... 41
                 .   :::|||      :::|||::::::
     216 TTCTAAAAGAAGAACTCACTGCCTGCTCCTCTCACATCTCCCTCTTTTCT 265

      42 ................LysAspLysIleGlyAspHisHisThrHisGlu 52
                         |||      |||   :::|||   :::|||:::
     266 TTTCTTCATCAACAAACAAGACACTCATTCAACAACATGCATCGCACCAA 315

      53 HisAsp...GluSerValSerHisValGlyLeuGln.............. 63
         |||:::   :::   :::   |||:::::::|||:::
     316 CATTCTTGCCAGCTCACACCACACATCGCTCTCAAACCCTCATCCCTCTT 365

      64 .......AlaHisPheGluThrTrpLeu......GlnMetHisHisAlaT 76
                ::::|||::::::|||   |||      :::   ||||||   |
     366 CACTCATTCTCACATCTCCACAACTCTCCCTCCTCACACTCACCATCTCA 415

      76 hrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGly 92
         ||      :::         :::   ::::::      :::
     416 CCCCGCCACAG.........CATGCCTCTCATGCCACCACCACCACTCAA 456

      93 AsnTyrLeuProProMetSerGlnLeuLeu.................Th 103
         :::   :::   |||   |||   ||||||
     457 CATCTCGTCTCTCCCAGCTCACCACTGCTCCTCTCTCACAACACTGCCAT 506

     103 rThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProP 120
          |||      |||   |||      :::|||   |||   :::|||||||
     507 CACCCACCTCTCATCACAGCTTCATGCTCACACACCCACCTCACTCCCAC 556

     120 roGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAsp 136
         ||:::|||
     557 CTCAACAC.......................................... 564

     137 LeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTy 153
                  |||:::::::||||||           :::|||   :::::
     565 .........CGCGCCTTTCTCCCA......CAACCGCTGCGCCTGCTCCA 599

     153 rArgAsnPro 156
```

```
        ::::||||||
    600 CCACAATCCA 609

seq_name: gb_est2:BI993724

seq_documentation_block:
LOCUS       BI993724               485 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  1031003D12.x2 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI993724
VERSION     BI993724.1  GI:16428451
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      145 a    151 c    105 g     84 t
ORIGIN


alignment_scores:
           Quality:   88.50            Length:     84
             Ratio:  1.735              Gaps:      5
 Percent Similarity:  60.714   Percent Identity: 33.333

alignment_block:
 US-09-674-779-2 x BI993724   ..

 Align seg 1/1  to: BI993724  from: 1  to: 485

      60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
         :::|||   |||||||||||||:::|||      |||::::::   |||::
     185 ATTGGCCAACAAGCACATTTCGACACA......CAGCAAAACGCAGCAGC 228

      76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
         ::::|||:::::::::|||   ::::::      :::::::|||   |||:

     229 AGAACAACACGCTATCCGCCAAGACACA......TCAAATCGCGCTGGCA 272

      93 snTyrLeuProProMetSerGlnLeuLeuThr...ThrAlaArgSerTrp 108
         ::            |||   |||   ||||||         |||||||||
     273 CC......AATCAAATGCACCAAAAACTCACAAAAAGACGGCGAAGCTGG 316

     109 GlnAlaCysGly......HisGluProTyrGlnLeuProProGluHisLe 123
         :::   |||:::      |||:::|||   |||      |||:::
     317 CGAAACTGCGCTCAAGTTCACAGGCCAGCGCAAGCGTGCCCACATGTGCG 366

     123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
          |||               :::   |||:::|||:::   :::   :::|
     367 CTGGTGCACGCGCTATCGGAACAAGCACGTGTACCGTGCTGTCGCCGACA 416

     140 rg 140
         ||
     417 GA 418

seq_name: gb_est2:BF177880

seq_documentation_block:
LOCUS       BF177880               509 bp    mRNA    linear   EST 31-OCT-2000
DEFINITION  Ljirnpest36-887-d7 Ljirnp Lambda HybriZap two-hybrid library Lotus
            japonicus cDNA clone LP887-36-d7 5' similar to mRNA for probable
            arabinogalactan-protein, mRNA sequence.
ACCESSION   BF177880
VERSION     BF177880.1  GI:11065571
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Poulsen,C. and Poedenphant,L.
  TITLE     Expressed sequence tags from Mesorhizobium loti infected roots of
            Lotus japonicus
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Poulsen,C.; Poedenphant,L.
            Laboratory of Gene Expression, Department of Molecular and
            Structural Biology
            University of Aarhus
            Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark
            Tel: +45 89425007
            Fax: +45 86201222
            Email: CHP@mbio.aau.dk
            Seq primer: pADGAl4 5'-primer CCA CTA CAA TGG ATG ATG TAT ATA AC
            High quality sequence stop: 518.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Lotus japonicus"
                     /cultivar="variety Gifu B-129"
                     /db_xref="taxon:34305"
                     /clone="LP887-36-d7"
                     /clone_lib="Ljirnp Lambda HybriZap two-hybrid library"
                     /dev_stage="young plants"
                     /lab_host="Escherichia coli XL-1Blue MRF'"
                     /note="Organ: whole roots and root nodule primordia;
                     Vector: pADGAL4; Site_1: EcoRI; Site_2: XhoI; Two times
                     polyA+ purified mRNA, oligodT primed, 5' adapter tag
                     GAATTCGGCACGAG "
BASE COUNT      100 a    221 c     92 g     96 t
ORIGIN


alignment_scores:
           Quality:   88.50            Length:    173
             Ratio:  1.196              Gaps:      9
 Percent Similarity:  42.775   Percent Identity: 27.746

alignment_block:
 US-09-674-779-2 x BF177880   ..
```

```
Align seg 1/1  to: BF177880  from: 1  to: 509

     49 HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHi 65
     97 CACCAACACAAGCACCGACAACAACCCCACCACCACCGCCAGCAGCAGCA 146

     65 sPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValA 82
    147 CCTGCCCCACCTCCAGCAACTCCACCAC..................... 174

     82 rgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMet 98
    175 ..CAGCAGCTACCCCTGCTCTCACCACTACTCCACCAGCTGCCACCCCTG 222

     99 SerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluPr 115
    223 .................................CTCCCTCCGCCTCC 236

    115 oTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuH 132
    237 CCTCCAGCTCCCACTCCAACCGCTTCCCCCACCGGTGCACCCACCCCTTC 286

    132 isLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGln 148
    287 CGCTT......CCTCTCCACCCGCCCCAATCCCATCCGGCCCAGCATCAG 330

    149 IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAl 165
    331 GCCCAGGCC............CCGCCGCTGGACCAG...GCCCAAACAGC 365

    165 aAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProA 182
    366 GCTGATACACCACCACCTC.........................CCTC 388

    182 spLeuGlu.IleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCy 198
    389 CGCTGCATTCTC...ACTCAACCAGCCCATCATTGCTGGCACCGCTCTTG 435

    198 sGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
    436 TTGCTACCTTCTTCGCCATGGT............CTTGGCTTAGGATTCT 473

    215 yrAlaThrGlyAlaIle 220
    474 TCCACGTGGAGGCCTTG 490

seq_name: gb_est2:BM002719

seq_documentation_block:
LOCUS       BM002719                523 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  1031106C05.x1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BM002719
VERSION     BM002719.1  GI:16437499
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      156 a    163 c    112 g     92 t
ORIGIN


alignment_scores:
            Quality:   88.50             Length:     84
              Ratio:  1.735               Gaps:      5
 Percent Similarity:  60.714   Percent Identity: 33.333

alignment_block:
 US-09-674-779-2 x BM002719   ..

 Align seg 1/1  to: BM002719  from: 1  to: 523

     60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
    183 ATTGGCCAACAAGCACATTTCGACACA......CAGCAAAACGCAGCAGC 226

     76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
    227 AGAACAACACGCTATCCGCCAAGACACA......TCAAATCGCGCTGGCA 270

     93 snTyrLeuProProMetSerGlnLeuLeuThr...ThrAlaArgSerTrp 108
    271 CC......AATCAAATGCACCAAAAACTCACAAAAAGACGGCGAAGCTGG 314

    109 GlnAlaCysGly......HisGluProTyrGlnLeuProProGluHisLe 123
    315 CGAAACTGCGCTCAAGTTCACAGGCCAGCGCAAGCGTGCCCACATGTGCG 364

    123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
    365 CTGGTGCACGCCCTATCGGAACAAGCACGTGTACCGTGCTGTCGCCGACA 414

    140 rg 140
    415 GA 416

seq_name: gb_est1:AW757658

seq_documentation_block:
LOCUS       AW757658                546 bp    mRNA    linear   EST 03-MAY-2000
DEFINITION  874002G08.x1 C. reinhardtii CC-1690, Lambda Zap II Chlamydomonas
            reinhardtii cDNA, mRNA sequence.
ACCESSION   AW757658
VERSION     AW757658.1  GI:7687010
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
```

```
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage."
BASE COUNT      160 a    168 c    126 g     92 t
ORIGIN


alignment_scores:
            Quality:   88.50            Length:     84
              Ratio:  1.735              Gaps:      5
 Percent Similarity:  60.714   Percent Identity: 33.333

alignment_block:
 US-09-674-779-2 x AW757658   ..

 Align seg 1/1  to: AW757658  from: 1  to: 546

      60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
         :::|||   |||||||||||||:::|||      |||:::::::   |||::
     102 ATTGGCCAACAAGCACATTTCGACACA......CAGCAAAACGCAGCAGC 145

      76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
         ::::|||:::::::::|||   ::::::      ::::::|||   |||:
     146 AGAACAACACGCTATCCGCCAAGACACA......TCAAATCGCGCTGGCA 189

      93 snTyrLeuProProMetSerGlnLeuLeuThr...ThrAlaArgSerTrp 108
         ::           |||   |||   ||||||         |||||||||
     190 CC......AATCAAATGCACCAAAAACTCACAAAAAGACGGCGAAGCTGG 233

     109 GlnAlaCysGly......HisGluProTyrGlnLeuProProGluHisLe 123
         :::   |||:::      |||:::|||   |||      |||:::
     234 CGAAACTGCGCTCAAGTTCACAGGCCAGCGCAAGCGTGCCCACATGTGCG 283

     123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
          |||               :::   |||:::|||:::   :::   :::|
     284 CTGGTGCACGCGCTATCGGAACAAGCACGTGTACCGTGCTGTCGCCGACA 333

     140 rg 140
         ||
     334 GA 335

seq_name: gb_gss:AQ345896
```

```
seq_documentation_block:
LOCUS       AQ345896                 564 bp    DNA     linear   GSS 07-MAY-1999
DEFINITION  RPCI11-126I23.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-126I23,
            DNA sequence.
ACCESSION   AQ345896
VERSION     AQ345896.1  GI:4170792
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Homo sapiens"
                     /db_xref="GDB:7548214"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-126I23"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      120 a    127 c    157 g    160 t
ORIGIN


alignment_scores:
            Quality:   88.50            Length:    180
              Ratio:  1.029              Gaps:     11
 Percent Similarity:  47.778   Percent Identity: 28.333

alignment_block:
 US-09-674-779-2 x AQ345896/rev   ..

 Align seg 1/1  to reverse of: AQ345896  from: 1  to: 564

      93 AsnTyrLeuProProMetSerGlnLeuLeuThr..............Th 104
         :::|||||||||||||   |||||||||:::|||
     512 CACTACCTTCCTCCCACCTCCCAGCTCATTACGATAGGAAATCCACTCCA 463

     104 rAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPro.... 119
          ||||||:::   ::::::   |||          ||||||||||||
     462 GGCACGTGCAGCCAAGAGCACTGGGCTTCTTTACTATCAGCTCCCAGTCA 413

     120 .................ProGluHisLeuTrpGlyGln......IleVal 128
                           ||||||   :::   :::|||      ::::::
     412 AGGGCTCTGGTATCTTCCCAGAAGGGGTAGGTAGCCAAGCATTGCTCATC 363

     129 ProThrLeuHisLeuTyr......GlnAspLeu..............Ly 138
         |||   :::   :::|||      |||   |||              ||
     362 CCTCTCTTTTCCATGTACACATTGCAGAAATTAATTTCCAGGAGAGGAAA 313
```

```
138 sSerArgGly...IleLeuProAlaAsnThrGlnIleArgSerValTyrA 154
    |:::||||||      |||||||||   ||||||
312 GACGAGAGGTCAGAGACTTCCTGCTATTACCCAGTC.............. 277

154 rgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHis 170
           |||:::::::   ||||||      |||   :::||||||
276 ..........CTCACTCATAGGGCAGGGGACCTGGCTCGAGAGAAGCAT 238

171 LeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSe 187
       ::::::   |||   :::            |||      :::
237 GCTGAGGATGCTAGGATCCCACTC.........ATCCTCACTCACAGAGT 197

187 rGlnAlaLeuTyrGlu...LeuGlnAsnArgLeuCysGlnTyrTrpLeuG 203
     :::||||||::::::      |||:::|||   |||:::
196 AGAAGCTCTACACCAGAAAAGGCAAAGCAGGAAGTGCAAA.......... 157

203 luHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAla 219
                                        ||||||   :::
156 ................................GCTGCTACCCCTGTC 142

220 IleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnPheSerGl 236
       |||   :::|||:::                  |||::::::|||
141 TGCCACCCTGAAACTAAA.................GCAGAGTTGTCACT 110

236 uThrAsnSerIleCysArgHisValLeuProLysAsnLys 249
     ::::::|||           :::   |||::::::::::
109 TAGTGAGAGCTGCACCACCATCCTCACCCCCAGCTCAAGA 70
```

OM of: US-09-674-779-2 to: N_Geneseq_032802:* out_format : pfs

Date: Sep 16, 2002 8:00 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09674779/runat_12092002_124144_26926/app_query.fasta_1.310
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=oli.rng -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-779-2
Query length: 250
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 199.450000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

| Sequence | Strd | Orig | ZScore | EScore | Len | ! Documentation .. |
|---|---|---|---|---|---|---|
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59797 | + | 250.00 | 4727.15 | 5.9e-2 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF28530 | + | 250.00 | 4701.71 | 1.5e-2 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI93830 | + | 9.00 | 153.44 | 3.37 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59798 | + | 8.00 | 152.79 | 3.66 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59799 | - | 8.00 | 149.78 | 5.39 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA07054 | + | 8.00 | 135.32 | 34.41 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK88795 | + | 8.00 | 135.32 | 34.41 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:AAQ10230 | - | 8.00 | 127.18 | 97.80 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAD08548 | - | 8.00 | 123.74 | 152.07 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH17512 | - | 8.00 | 123.48 | 157.11 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:AAQ10484 | - | 8.00 | 122.46 | 179.20 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS46451 | - | 8.00 | 115.26 | 450.98 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL18726 | + | 8.00 | 102.35 | 2.4e+0 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ01425 | - | 8.00 | 77.38 | 5.8e+04 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL31302 | - | 7.00 | 131.99 | 52.76 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA67992 | - | 7.00 | 128.11 | 86.74 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI22900 | - | 7.00 | 128.11 | 86.74 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA18813 | - | 7.00 | 124.55 | 136.97 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC05959 | - | 7.00 | 123.24 | 162.17 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX51506 | - | 7.00 | 123.10 | 164.96 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS65817 | + | 7.00 | 123.02 | 166.82 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC14154 | - | 7.00 | 121.56 | 200.99 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA73240 | + | 7.00 | 120.95 | 217.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK21676 | + | 7.00 | 120.95 | 217.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK47841 | + | 7.00 | 120.95 | 217.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI53670 | + | 7.00 | 120.95 | 217.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL12381 | - | 7.00 | 119.85 | 250.35 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL21257 | - | 7.00 | 119.85 | 250.35 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK65002 | + | 7.00 | 119.79 | 252.17 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAV89764 | - | 7.00 | 119.74 | 253.99 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA50979 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA68959 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA35907 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK17286 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK43076 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI23842 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI49154 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI09449 | + | 7.00 | 119.30 | 268.51 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS61584 | + | 7.00 | 119.23 | 271. | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA11586 | - | 7.00 | 118.97 | 280 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV37035 | + | 7.00 | 118.55 | 295. | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH01713 | + | 7.00 | 118.55 | 295 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT16270 | + | 7.00 | 118.41 | 301. | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH01012 | + | 7.00 | 118.09 | 313 | | |
| /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH01298 | + | 7.00 | 118.00 | 317 | | |

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59797

seq_documentation_block:
ID AAF59797 standard; DNA; 753 BP.
XX
AC AAF59797;
XX
DT 04-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis strain ATCC43617 BASB120 DNA.
XX
KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory; ds.
XX
OS Moraxella catarrhalis.
XX
PN WO200109335-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07361.
XX
PR 03-AUG-1999; 99GB-0018281.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2001-159872/16.
DR P-PSDB; AAB60645.
XX
PT New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
PT strain American Type Culture Collection 43617, for use as therapeutic
PT agents or vaccines against bacterial infections, e.g. otitis media or
PT pneumonia -
XX
PS Claim 13; Page 64; 75pp; English.
XX
CC The invention relates to the Moraxella catarrhalis strain ATCC43617
CC BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
CC invention also relates to immunogenic fragments of the BASB120 protein,
CC expression vectors and host cells comprising BASB120 nucleic acids, the
CC recombinant production of BASB120, vaccine compositions comprising the
CC BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
CC compositions comprising the anti-BASB120 antibody, and a method of
CC identifying a Moraxella catarrhalis infection via the detection of
CC BASB120 proteins or antibodies. The vaccine compositions of the invention
CC are useful as prophylactic or therapeutic agents against Moraxella
CC catarrhalis infections in mammals, particularly humans. Moraxella
CC catarrhalis is a Gram negative bacterium frequently isolated from the
CC human upper respiratory tract, which is responsible for several
CC pathological conditions. It is responsible for about 15% of otitis media
CC cases in children (which can lead to temporary or permanent hearing
CC loss). It also causes pneumonia in elderly people, and sinusitis,
CC nosocomial infections and, less frequently, invasive diseases. BASB120
CC proteins or nucleotides may additionally be used in screening for novel
CC antibacterial compounds, and in the diagnosis and staging of infections.
CC The present sequence represents DNA encoding the Moraxella catarrhalis
CC strain ATCC43617 BASB120 protein.
XX
SQ Sequence 753 BP; 239 A; 166 C; 152 G; 196 T; 0 other;

```
alignment_scores:
            Quality:  250.00              Length:    250
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF59797   ..

 Align seg 1/1  to: AAF59797  from: 1  to: 753

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 50

      17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 100

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 150

      51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 200

      67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 250

      84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 300

     101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     301 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 350

     117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 400

     134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 450

     151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     451 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 500

     167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     501 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 550

     184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     551 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 600

     201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     601 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 650

     217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     651 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 700

     234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 750

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF28530

seq_documentation_block:

ID   AAF28530 standard; DNA; 23210 BP.
XX
AC   AAF28530;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #17.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 141-146; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 23210 BP; 7108 A; 4482 C; 4951 G; 6668 T; 1 other;


alignment_scores:
            Quality:  250.00              Length:    250
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF28530   ..

 Align seg 1/1  to: AAF28530  from: 1  to: 23210

       1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   11483 ATGAAAAATTTTAATCAATACTTTATAACTACACTTATCAGCAGTATGCT 11532

      17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   11533 GGTTGCATGCAGTGCACCCATACCAACCAATCCACAAGTATCCCCAATAA 11582

      34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   11583 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATCATACA 11632

      51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   11633 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTTGCAGGCACATTTTGA 11682
```

```
     67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11683 GACTTGGCTACAGATGCACCATGCCACCAAACAAGAGGTAGTTAGGTATC 11732

     84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11733 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 11782

    101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11783 CTACTAACGACTGCACGGTCATGGCAGGCATGTGGTCATGAACCTTATCA 11832

    117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11833 GCTGCCACCAGAACATCTTTGGGGTCAGATTGTACCAACATTGCACTTGT 11882

    134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11883 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAATACCCAAATTCGC 11932

    151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11933 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 11982

    167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  11983 GAGTAAGCATTTGACCAATAGTGCCATTGATATTTGGGTGCCTGACCTTG 12032

    184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12033 AAATAAAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 12082

    201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12083 TGGCTAGAGCATGGCGAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 12132

    217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12133 AGGGGCGATTCATCTGGATACCCAAGGGTTTAGAAAATGGGGTGCTCAAT 12182

    234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  12183 TTTCTGAAACAAACTCTATTTGTCGTCATGTCTTACCAAAAAATAAGCTA 12232
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI93830

seq_documentation_block:
ID AAI93830 standard; cDNA; 474 BP.
XX
AC AAI93830;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 13890.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO13899.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 13890; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 474 BP; 116 A; 109 C; 139 G; 110 T; 0 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAI93830 ..

Align seg 1/1 to: AAI93830 from: 1 to: 474

```
     33 IleLysThrProSerValLeuIleThr 41
        |||||||||||||||||||||||||||
    359 ATCAAGACCCCATCTGTACTAATAACG 385
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59798

seq_documentation_block:
ID AAF59798 standard; DNA; 40 BP.
XX
AC AAF59798;
XX
DT 04-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis BASB120 mutagenic PCR primer, SEQ ID NO:3.
XX
KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory; mutagenic PCR primer; ss.
XX
OS Moraxella catarrhalis.
OS Synthetic.
XX
PN WO200109335-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07361.
XX
PR 03-AUG-1999; 99GB-0018281.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```
PI   Thonnard J;
XX
DR   WPI; 2001-159872/16.
XX
PT   New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
PT   strain American Type Culture Collection 43617, for use as therapeutic
PT   agents or vaccines against bacterial infections, e.g. otitis media or
PT   pneumonia -
XX
PS   Example 2; Page 51; 75pp; English.
XX
CC   The invention relates to the Moraxella catarrhalis strain ATCC43617
CC   BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
CC   invention also relates to immunogenic fragments of the BASB120 protein,
CC   expression vectors and host cells comprising BASB120 nucleic acids, the
CC   recombinant production of BASB120, vaccine compositions comprising the
CC   BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
CC   compositions comprising the anti-BASB120 antibody, and a method of
CC   identifying a Moraxella  catarrhalis infection via the detection of
CC   BASB120 proteins or antibodies. The vaccine compositions of the invention
CC   are useful as prophylactic or therapeutic agents against Moraxella
CC   catarrhalis infections in mammals, particularly humans. Moraxella
CC   catarrhalis is a Gram negative bacterium frequently isolated from the
CC   human upper respiratory tract, which is responsible for several
CC   pathological conditions. It is responsible for about 15% of otitis media
CC   cases in children (which can lead to temporary or permanent hearing
CC   loss). It also causes pneumonia in elderly people, and sinusitis,
CC   nosocomial infections and, less frequently, invasive diseases. BASB120
CC   proteins or nucleotides may additionally be used in screening for novel
CC   antibacterial compounds, and in the diagnosis and staging of infections.
CC   Sequences AAF59798-AAF59799 represent PCR primers used to introduce
CC   EcoRI and SalI restriction sites into Moraxella catarrhalis strain
CC   ATCC43617 BASB120 DNA to facilitate cloning into the pTLZ2 vector.
XX
SQ   Sequence 40 BP; 17 A; 4 C; 7 G; 12 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF59798   ..

 Align seg 1/1  to: AAF59798  from: 1  to: 40

       1 MetLysAsnPheAsnGlnTyrPhe 8
         ||||||||||||||||||||||||
      16 ATGAAAAATTTTAATCAATACTTT 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAF59799

seq_documentation_block:
ID   AAF59799 standard; DNA; 60 BP.
XX
AC   AAF59799;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Moraxella catarrhalis BASB120 mutagenic PCR primer, SEQ ID NO:4.
XX
KW   BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
KW   genetic immunisation; infection; upper respiratory tract; otitis media;
KW   hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW   invasive disease; antibacterial; auditory; mutagenic PCR primer; ss.
XX
OS   Moraxella catarrhalis.
OS   Synthetic.
XX
PN   WO200109335-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07361.
XX
PR   03-AUG-1999;   99GB-0018281.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2001-159872/16.
XX
PT   New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
PT   strain American Type Culture Collection 43617, for use as therapeutic
PT   agents or vaccines against bacterial infections, e.g. otitis media or
PT   pneumonia -
XX
PS   Example 2; Page 51; 75pp; English.
XX
CC   The invention relates to the Moraxella catarrhalis strain ATCC43617
CC   BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
CC   invention also relates to immunogenic fragments of the BASB120 protein,
CC   expression vectors and host cells comprising BASB120 nucleic acids, the
CC   recombinant production of BASB120, vaccine compositions comprising the
CC   BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
CC   compositions comprising the anti-BASB120 antibody, and a method of
CC   identifying a Moraxella  catarrhalis infection via the detection of
CC   BASB120 proteins or antibodies. The vaccine compositions of the invention
CC   are useful as prophylactic or therapeutic agents against Moraxella
CC   catarrhalis infections in mammals, particularly humans. Moraxella
CC   catarrhalis is a Gram negative bacterium frequently isolated from the
CC   human upper respiratory tract, which is responsible for several
CC   pathological conditions. It is responsible for about 15% of otitis media
CC   cases in children (which can lead to temporary or permanent hearing
CC   loss). It also causes pneumonia in elderly people, and sinusitis,
CC   nosocomial infections and, less frequently, invasive diseases. BASB120
CC   proteins or nucleotides may additionally be used in screening for novel
CC   antibacterial compounds, and in the diagnosis and staging of infections.
CC   Sequences AAF59798-AAF59799 represent PCR primers used to introduce
CC   EcoRI and SalI restriction sites into Moraxella catarrhalis strain
CC   ATCC43617 BASB120 DNA to facilitate cloning into the pTLZ2 vector.
XX
SQ   Sequence 60 BP; 14 A; 5 C; 21 G; 20 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF59799/rev   ..

 Align seg 1/1  to reverse of: AAF59799  from: 1  to: 60

     243 HisValLeuProLysAsnLysLeu 250
         ||||||||||||||||||||||||
      60 CATGTCTTACCAAAAAATAAGCTA 37

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA07054

seq_documentation_block:
ID   ABA07054 standard; cDNA; 421 BP.
XX
AC   ABA07054;
XX
DT   14-JAN-2002  (first entry)
XX
DE   Human pancreatic cancer related cDNA, SEQ ID NO: 183.
XX
KW   Human; cytostatic; antidiabetic; antiinflammatory; gastric; osteopathic;
KW   antihormone; antiulcer; thyroid-active; gene therapy; antisense therapy;
KW   pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis;
KW   diabetes; endocrine disorder; acromegaly; hyperthyroidism;
```

KW gastrointestinal disorder; Crohn's disease; duodenal ulcer; ss.
XX
OS Homo sapiens.
XX
PN WO200155206-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01353.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457717/49.
DR P-PSDB; ABB10724.
XX
PT Isolated pancreatic cancer polypeptide for treating, preventing and/ or
PT prognosing disorders related to the pancreas including pancreatic
PT cancers and also for testing and detection e.g. diagnosis -
XX
PS Claim 1; SEQ ID NO 183; 537pp; English.
XX
CC The invention relates to an isolated polypeptide comprising an amino
CC acid sequence at least 90% identical to 188 amino acid sequences fully
CC defined in the specification and encoded by 188 cDNA clones fully
CC defined in the specification. The invention also relates to a fragment
CC having biological activity, a domain, an epitope, full length protein,
CC variant, allelic variant or a species homologue of the fully defined
CC sequence. The polynucleotide and polypeptide are useful for treating,
CC preventing and/or prognosing disorders related to the pancreas including
CC pancreatic cancer, pancreatitis, diabetes, endocrine disorders such as
CC acromegaly or hyperthyroidism, and gastrointestinal disorders such as
CC Crohn's disease and duodenal ulcers. The present sequence is a
CC pancreatic cancer-related polynucleotide of the invention.
XX
SQ Sequence 421 BP; 124 A; 79 C; 91 G; 126 T; 1 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x ABA07054 ..

Align seg 1/1 to: ABA07054 from: 1 to: 421

211 GlyLeuGlyLeuTyrAlaThrGly 218
||||||||||||||||||||||||
384 GGTTTGGGGCTTTATGCCACTGGG 407

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK88795

seq_documentation_block:
ID AAK88795 standard; cDNA; 421 BP.
XX
AC AAK88795;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen coding sequence SEQ ID NO: 1111.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ss.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
DR P-PSDB; AAM93022.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Claim 1; SEQ ID NO 1111; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC system antigen of the invention.
XX
SQ Sequence 421 BP; 124 A; 79 C; 91 G; 126 T; 1 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAK88795 ..

Align seg 1/1 to: AAK88795 from: 1 to: 421

211 GlyLeuGlyLeuTyrAlaThrGly 218
||||||||||||||||||||||||
384 GGTTTGGGGCTTTATGCCACTGGG 407

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:AAQ10230

seq_documentation_block:
ID AAQ10230 standard; DNA; 1262 BP.
XX
AC AAQ10230;
XX
DT 28-MAR-1991 (first entry)
XX
DE Synthetic Streptokinase gene.
XX
KW streptokinase; thrombolytic agent; myocardial infarction; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 15..1256
FT /*tag= a
FT /product= streptokinase
XX
PN EP407942-A.
XX
PD 16-JAN-1991.
XX
PF 09-JUL-1990; 90EP-0113099.
XX
PR 11-APR-1990; 90JP-0096830.
PR 11-JUL-1989; 89JP-0179432.
PR 27-NOV-1989; 89JP-0307957.
XX
PA (SAKA ) OTSUKA PHARM FACTOR.
XX
PI Majima E, Ogino K, Ono K, Sakata Y, Uenoyama T;
XX
DR WPI; 1991-016179/03.
DR P-PSDB; AAR10194.
XX
PT Synthetic gene encoding streptokinase - scale, high purity prodn.
PT of streptokinase used as a thrombolytic agent
XX
PS Claim 3; Page 56; 76pp; English.
XX
CC The 5' end of the coding strand overhangs the 3' end of the
CC complementary strand by 4 bases; the 5' end of the complementary
CC strand overhangs the 3' end of the sense strand by AGCT. The
CC sequence encoding streptokinase was first divided into 52
CC oligonucleotide fragments of 43-56 bases each. The individual
CC fragments were synthesised by solid-phase beta-cyanomethyl
CC phosphoamidide method, phosphate groups added and the fragments
CC ligated to obtain pSKK. This was inserted into expression vector

```
CC   pKTN2-2 and used to transform E.coli JM109 which was cultured to
CC   produce streptokinase. See also AAR10195-R10200.
XX
SQ   Sequence 1262 BP; 365 A; 350 C; 262 G; 285 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAQ10230/rev   ..

 Align seg 1/1  to reverse of: AAQ10230  from: 1  to: 1262

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
     393 GAGTTACAGAACCGTCTTTGTCAG 370

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAD08548

seq_documentation_block:
ID   AAD08548 standard; DNA; 2007 BP.
XX
AC   AAD08548;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Feline calcivirus (FCV) capsid gene.
XX
KW   Raccoon poxvirus; RCNV; thymidine kinase; TK; vaccine; haemagglutinin;
KW   HA; feline pathogen; feline panleukopenia virus; FPV; feline calcivirus;
KW   FCV; capsid protein; virucide; ds.
XX
OS   Feline calcivirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2007
FT                   /*tag=  a
FT                   /product= "Feline calcivirus (FCV) capsid protein"
XX
PN   US6241989-B1.
XX
PD   05-JUN-2001.
XX
PF   03-NOV-1995;   95US-0552369.
XX
PR   09-JUL-1991;   91US-0726609.
PR   27-JAN-1994;   94US-0190789.
XX
PA   (CORR ) CORNELL RES FOUND INC.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Scott FW,  Ngichabe CK,  Hu L,  Esposito JJ;
XX
DR   WPI; 2001-407214/43.
DR   P-PSDB; AAE04304.
XX
PT   New multivalent recombinant raccoon poxviruses, useful as vaccines to
PT   immunize felines against subsequent challenge by feline pathogens  -
XX
PS   Example 6; Column 31-36; 35pp; English.
XX
CC   The present invention relates to multivalent recombinant raccoon
CC   poxviruses (RCNV), containing more than one exogenous gene inserted
CC   into either the thymidine kinase (TK) gene, the haemagglutinin (HA)
CC   gene, or a combination thereof. The multivalent recombinant raccoon
CC   poxviruses are useful as vaccines to immunise felines against subsequent
CC   challenge by feline pathogens. The recombinant multivalent vaccine is
CC   formed by inserting multiple genes such as a feline panleukopenia virus
CC   (FPV) gene, and/or a feline calcivirus (FCV) capsid protein gene each
CC   operably linked to a promoter, into a raccoon poxvirus for expression.
CC   The present sequence is Feline calcivirus (FCV) capsid protein encoding
CC   gene.
XX
SQ   Sequence 2007 BP; 528 A; 473 C; 421 G; 585 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAD08548/rev   ..

 Align seg 1/1  to reverse of: AAD08548  from: 1  to: 2007

     192 GluLeuGlnAsnArgLeuCysGln 199
         |||||||||||||||||||||||||
    1090 GAATTACAAAATCGGTTATGTCAG 1067

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH17512

seq_documentation_block:
ID   AAH17512 standard; cDNA; 2077 BP.
XX
AC   AAH17512;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:16985.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 16985; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
```

```
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2077 BP; 630 A; 452 C; 395 G; 600 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAH17512/rev   ..

 Align seg 1/1  to reverse of: AAH17512  from: 1  to: 2077

     211 GlyLeuGlyLeuTyrAlaThrGly 218
         ||||||||||||||||||||||||
     937 GGACTAGGACTTTATGCCACTGGA 914

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:AAQ10484

seq_documentation_block:
ID   AAQ10484 standard; cDNA; 2385 BP.
XX
AC   AAQ10484;
XX
DT   22-APR-1991  (first entry)
XX
DE   Encodes Feline calcivirus capsid protein.
XX
KW   feline calcivirus capsid protein; FCV; multivalent vaccine; cat; ds.
XX
OS   feline calcivirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             18..2024
FT                   /*tag=  a
FT                   /product= capsid protein
XX
PN   WO9101332-A.
XX
PD   07-FEB-1991.
XX
PF   09-JUL-1990;   90WO-US03753.
XX
PR   21-JUL-1989;   89US-0383909.
PR   18-SEP-1989;   89US-0408989.
XX
PA   (UPJO ) UPJOHN CO.
XX
PI   Wardley R,  Post LE;
XX
DR   WPI; 1991-058117/08.
DR   P-PSDB; AAR10686.
XX
PT   Feline calcivirus capsid protein - from recombinant DNA for subunit
PT   and multivalent vaccines for FCV
XX
PS   Claim 3; Page 28; 40pp; English.
XX
CC   CRFK cells were inoculated, total poly-A+ RNA was isolated and FCV
CC   double-stranded RNA was LiCl fractionated from this RNA. Double-
CC   stranded cDNA was prepared , blunt-ended with t4 DNA polymerase and
CC   treated with EcoRI methylase prior to ligation of EcoRI linkers. The
CC   mixture was digested with EcoRI and ligated to lambda gt10. E.coli
CC   strains transformed with the viral vectors were plated and positive
CC   plaques were transferred to nitrocellulose. They were probed with
CC   random-primed cDNA prepared from FCV genomic RNA. A plaque with a
CC   strong hybridisation signal was isolated and found to contain an
CC   EcoRI fragment (ca.4200bp). It was cloned in pUC18 to give pCV2.
CC   A second library was constructed in pUC18 using PstI linkers. A
CC   plasmid was selected from this library and designated pCV7. It was
CC   found to have restriction sites in common with pCV2. These two
CC   plasmids were used to isolate other regions of the FCV genome.
XX
SQ   Sequence 2385 BP; 642 A; 562 C; 502 G; 679 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAQ10484/rev   ..

 Align seg 1/1  to reverse of: AAQ10484  from: 1  to: 2385

     192 GluLeuGlnAsnArgLeuCysGln 199
         ||||||||||||||||||||||||
    1107 GAATTACAAAATCGGTTATGTCAG 1084

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS46451

seq_documentation_block:
ID   AAS46451 standard; DNA; 6291 BP.
XX
AC   AAS46451;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #173.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 173; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
```

```
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6291 BP; 1501 A; 210 C; 1626 G; 2954 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAS46451/rev   ..

 Align seg 1/1  to reverse of: AAS46451  from: 1  to: 6291

      34 LysThrProSerValLeuIleThr 41
         |||||||||||||||||||||||||
    1884 AAAACACCATCCGTTTTAATTACT 1861

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:ABL18726

seq_documentation_block:
ID   ABL18726 standard; DNA; 35832 BP.
XX
AC   ABL18726;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 7651.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions
XX
PS   Claim 1; SEQ ID NO 7651; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 35832 BP; 9547 A; 8198 C; 8102 G; 9985 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABL18726   ..

 Align seg 1/1  to: ABL18726  from: 1  to: 35832

       8 PheIleThrThrLeuIleSerSer 15
         |||||||||||||||||||||||
   20967 TTCATTACTACGCTTATAAGCTCT 20990

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAZ01425

seq_documentation_block:
ID   AAZ01425 standard; DNA; 1038602 BP.
XX
AC   AAZ01425;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Complete genome sequence of Chlamydia trachomatis.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Claim 1; Page 373-656; 1755pp; English.
XX
CC   The present sequence represents the complete genome of Chlamydia
CC   trachomatis. Open reading frames (ORFs) of the genome encode
CC   polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
CC   against Chlamydia trachomatis. Antisense and ribozyme sequences can also
```

```
CC   be used to control growth of the microorganism. Chlamydia trachomatis is
CC   responsible for a large number of diseases, e.g. eye diseases such as
CC   conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC   conjunctivitis; genital diseases such as nongonococcal uretritis,
CC   epidymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
CC   pneumopathy in breast feeding infants; and venereal
CC   lymphogranulomatosis. The polypeptides of the invention may be of use in
CC   treating these diseases.
XX
SQ   Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAZ01425/rev   ..

 Align seg 1/1  to reverse of: AAZ01425  from: 1  to: 1038602

     212 LeuGlyLeuTyrAlaThrGlyAla 219
         ||||||||||||||||||||||||
  550259 CTAGGGTTATACGCAACAGGAGCG 550236

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL31302

seq_documentation_block:
ID   AAL31302 standard; DNA; 51 BP.
XX
AC   AAL31302;
XX
DT   24-JAN-2002  (first entry)
XX
DE   Human SNP oligonucleotide #4510.
XX
KW   Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
KW   neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
KW   amyloid protein; angiopoietin; apoptosis related protein; cadherin;
KW   cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
KW   complement related protein; cytochrome; kinesin; cytokine; interferon;
KW   interleukin; G-protein coupled receptor; thioesterase; inflammation;
KW   multifactorial disease; autoimmune disease; infection;
KW   nervous system disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200147944-A2.
XX
PD   05-JUL-2001.
XX
PF   28-DEC-2000; 2000WO-US35498.
XX
PR   28-DEC-1999;   99US-0173419.
PR   27-DEC-2000; 2000US-0173419.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2001-465210/50.
XX
PT   Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
PT   oncogenes and histones, useful for diagnosing and treating, e.g.
PT   cancer, autoimmune diseases and infections -
XX
PS   Claim 1; Page 2683; 4143pp; English.
XX
CC   The present invention relates to oligonucleotides encoding polymorphic
CC   variants of proteins related to amylases, amyloid proteins, angiopoietin,
CC   apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
CC   histones, kinases, colony stimulating factors, complement related
CC   proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
CC   G-protein coupled receptors and thioesterases. The present sequence is
CC   one such oligonucleotide. The oligonucleotides and the peptides encoded
CC   by them may be used in the prevention, diagnosis and treatment of
CC   diseases associated with inappropriate expression of the proteins listed
CC   above. Disorders that may be prevented, diagnosed and/or treated include
CC   multifactorial diseases with a genetic component, such as autoimmune
CC   diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
CC   systemic lupus erythromatosus and Grave's disease), inflammation, cancer
CC   (e.g. cancers of the bladder, brain, breast, colon and kidney,
CC   leukaemia), diseases of the nervous system and an infection of pathogenic
CC   organisms.
XX
SQ   Sequence 51 BP; 10 A; 9 C; 16 G; 16 T; 0 other;


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAL31302/rev   ..

 Align seg 1/1  to reverse of: AAL31302  from: 1  to: 51

      19 AlaCysSerAlaProIlePro 25
         |||||||||||||||||||||
      50 GCATGCTCAGCCCCCATCCCA 30

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA67992

seq_documentation_block:
ID   ABA67992 standard; DNA; 86 BP.
XX
AC   ABA67992;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #16297.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 4; SEQ ID NO 16297; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
```

```
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 86 BP; 24 A; 12 C; 11 G; 39 T; 0 other;


alignment_scores:
           Quality:    7.00             Length:      7
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA67992/rev   ..

 Align seg 1/1  to reverse of: ABA67992  from: 1  to: 86

      98 MetSerGlnLeuLeuThrThr 104
         ||||||||||||||||||||
      53 ATGTCACAACTTTTGACTACA 33

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI22900

seq_documentation_block:
ID   AAI22900 standard; DNA; 86 BP.
XX
AC   AAI22900;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #12833 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 12833; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 86 BP; 24 A; 12 C; 11 G; 39 T; 0 other;


alignment_scores:
           Quality:    7.00             Length:      7
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAI22900/rev   ..

 Align seg 1/1  to reverse of: AAI22900  from: 1  to: 86

      98 MetSerGlnLeuLeuThrThr 104
         ||||||||||||||||||||
      53 ATGTCACAACTTTTGACTACA 33

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA18813

seq_documentation_block:
ID   ABA18813 standard; DNA; 139 BP.
XX
AC   ABA18813;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11144.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
```

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 11144; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 139 BP; 31 A; 27 C; 24 G; 57 T; 0 other;

alignment_scores:

```
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA18813/rev   ..

 Align seg 1/1  to reverse of: ABA18813  from: 1  to: 139

      17 LeuValAlaCysSerAlaPro 23
         ||||||||||||||||||||
     109 TTAGTGGCGTGCAGTGCACCA 89

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC05959

seq_documentation_block:
ID   AAC05959 standard; cDNA; 166 BP.
XX
AC   AAC05959;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 10034.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 10034; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 166 BP; 41 A; 35 C; 36 G; 51 T; 3 other;


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAC05959/rev   ..

 Align seg 1/1  to reverse of: AAC05959  from: 1  to: 166

     141 GlyIleLeuProAlaAsnThr 147
         ||||||||||||||||||||
     122 GGCATACTGCCTGCAAACACA 102

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAX51506

seq_documentation_block:
ID   AAX51506 standard; cDNA; 169 BP.
XX
AC   AAX51506;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:85.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906549-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01231.
XX
PR   01-AUG-1997;   97US-0905279.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153779/13.
DR   P-PSDB; AAY12728.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS   Claim 1; Page 208; 522pp; English.
XX
CC   AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12681 to
CC   AAY12913, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 169 BP; 42 A; 35 C; 38 G; 51 T; 3 other;


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-674-779-2 x AAX51506/rev   ..

 Align seg 1/1  to reverse of: AAX51506  from: 1  to: 169

     141 GlyIleLeuProAlaAsnThr 147
         ||||||||||||||||||||
     123 GGCATACTGCCTGCAAACACA 103

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:AAS65817

seq_documentation_block:
ID   AAS65817 standard; cDNA; 171 BP.
XX
AC   AAS65817;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #1621.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG01630.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 1621; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 171 BP; 46 A; 39 C; 59 G; 27 T; 0 other;

alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAS65817   ..

 Align seg 1/1  to: AAS65817  from: 1  to: 171

     162 AlaGlyGlyAlaAlaMetSer 168
         ||||||||||||||||||||
     115 GCAGGAGGGGCAGCCATGAGT 135

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC14154

seq_documentation_block:
ID   AAC14154 standard; cDNA; 208 BP.
XX
AC   AAC14154;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 18229.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 18229; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 208 BP; 62 A; 35 C; 56 G; 54 T; 1 other;

alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-674-779-2 x AAC14154/rev   ..

 Align seg 1/1  to reverse of: AAC14154  from: 1  to: 208

      30 ValSerProIleLysThrPro 36
         ||||||||||||||||||||
     126 GTGAGCCCAATTAAAACTCCT 106

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA73240

seq_documentation_block:
ID   ABA73240 standard; DNA; 226 BP.
XX
AC   ABA73240;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #21545.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 4; SEQ ID NO 21545; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA73240   ..

 Align seg 1/1  to: ABA73240  from: 1  to: 226

      35 ThrProSerValLeuIleThr 41
         ||||||||||||||||||||
     171 ACACCATCAGTATTAATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK21676

seq_documentation_block:
ID   AAK21676 standard; DNA; 226 BP.
XX
AC   AAK21676;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 21667.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 21667; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAK21676   ..

 Align seg 1/1  to: AAK21676  from: 1  to: 226

      35 ThrProSerValLeuIleThr 41
         ||||||||||||||||||||
     171 ACACCATCAGTATTAATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK47841
```

```
seq_documentation_block:
ID   AAK47841 standard; DNA; 226 BP.
XX
AC   AAK47841;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 22398.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 22398; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAK47841   ..

 Align seg 1/1  to: AAK47841  from: 1  to: 226

      35 ThrProSerValLeuIleThr 41
         |||||||||||||||||||||
     171 ACACCATCAGTATTAATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI53670

seq_documentation_block:
ID   AAI53670 standard; DNA; 226 BP.
XX
AC   AAI53670;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #22356 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 22356; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAI53670   ..

 Align seg 1/1  to: AAI53670  from: 1  to: 226

      35 ThrProSerValLeuIleThr 41
         |||||||||||||||||||||
     171 ACACCATCAGTATTAATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL12381

seq_documentation_block:
ID   AAL12381 standard; cDNA; 262 BP.
XX
AC   AAL12381;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 4838.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
```

```
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 869; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 262 BP; 70 A; 63 C; 64 G; 65 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAL12381/rev   ..

 Align seg 1/1  to reverse of: AAL12381  from: 1  to: 262

     159 AsnGlnCysAlaGlyGlyAla 165
         ||||||||||||||||||||
      76 AACCAGTGCGCAGGGGGTGCA 56

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAL21257

seq_documentation_block:
ID   AAL21257 standard; cDNA; 262 BP.
XX
AC   AAL21257;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 13714.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 2439; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 262 BP; 71 A; 63 C; 63 G; 65 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAL21257/rev   ..

 Align seg 1/1  to reverse of: AAL21257  from: 1  to: 262

     159 AsnGlnCysAlaGlyGlyAla 165
         ||||||||||||||||||||
      76 AACCAGTGCGCAGGGGGTGCA 56

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK65002

seq_documentation_block:
ID   AAK65002 standard; DNA; 264 BP.
XX
AC   AAK65002;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19814.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
```

PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 19814; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to

```
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 264 BP; 73 A; 50 C; 46 G; 95 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAK65002   ..

 Align seg 1/1  to: AAK65002  from: 1  to: 264

     164 GlyAlaAlaMetSerLysHis 170
         ||||||||||||||||||||
      33 GGAGCAGCAATGTCTAAACAT 53

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:AAV89764

seq_documentation_block:
ID   AAV89764 standard; cDNA; 266 BP.
XX
AC   AAV89764;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone CT159.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 322; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 266 BP; 79 A; 38 C; 54 G; 95 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAV89764/rev   ..

 Align seg 1/1  to reverse of: AAV89764  from: 1  to: 266

      12 LeuIleSerSerMetLeuVal 18
         ||||||||||||||||||||
      87 TTAATCTCTTCTATGTTGGTC 67

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA50979

seq_documentation_block:
ID   ABA50979 standard; DNA; 282 BP.
XX
AC   ABA50979;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human breast cell single exon nucleic acid probe #9674.
XX
KW   Human; microarray; single exon probe; gene expression; breast;
KW   disease; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00662.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes,
PT   useful for measuring gene expression in sample derived from human
PT   breast, comprises number of single exon nucleic acid probes
XX
PS   Claim 4; SEQ ID NO 9674; 327pp + sequence listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting
CC   the probes with a collection of detectably labelled nucleic acids
CC   derived from mRNA of human breast, and then measuring the label
CC   bound to each probe of the microarray. The probes are useful for
CC   verifying the expression of regions of genomic DNA predicted to
CC   encode proteins. They are useful for gene discovery, and for
```

```
CC   determining predisposition and/or prognosing breast disease. Gene
CC   expression analysis is useful for assessing the toxicity of chemical
CC   agents on cells. The microarray of this invention presents a far greater
CC   diversity of probes for measuring gene expression, with far less bias
CC   than expressed sequence tag microarrays. The method is suitable for
CC   rapid production of functional information from genomic sequence. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
           Quality:    7.00            Length:      7
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA50979   ..

 Align seg 1/1  to: ABA50979  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA68959

seq_documentation_block:
ID   ABA68959 standard; DNA; 282 BP.
XX
AC   ABA68959;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #17264.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 4; SEQ ID NO 17264; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
           Quality:    7.00            Length:      7
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA68959   ..

 Align seg 1/1  to: ABA68959  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA35907

seq_documentation_block:
ID   ABA35907 standard; DNA; 282 BP.
XX
AC   ABA35907;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #14373 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 4; SEQ ID No 14373; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease.
```

```
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA35907   ..

 Align seg 1/1  to: ABA35907  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK17286

seq_documentation_block:
ID   AAK17286 standard; DNA; 282 BP.
XX
AC   AAK17286;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 17277.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 17277; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAK17286   ..

 Align seg 1/1  to: AAK17286  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK43076

seq_documentation_block:
ID   AAK43076 standard; DNA; 282 BP.
XX
AC   AAK43076;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 17633.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 17633; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAK43076   ..
```

```
 Align seg 1/1  to: AAK43076  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI23842

seq_documentation_block:
ID   AAI23842 standard; DNA; 282 BP.
XX
AC   AAI23842;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #13775 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 13775; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAI23842   ..

 Align seg 1/1  to: AAI23842  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI49154

seq_documentation_block:
ID   AAI49154 standard; DNA; 282 BP.
XX
AC   AAI49154;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #17840 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 17840; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAI49154   ..

 Align seg 1/1  to: AAI49154  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI09449

seq_documentation_block:
ID   AAI09449 standard; DNA; 282 BP.
XX
AC   AAI09449;
```

```
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #9440 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 9440; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAI09449   ..

 Align seg 1/1  to: AAI09449  from: 1  to: 282

     100 GlnLeuLeuThrThrAlaArg 106
         ||||||||||||||||||||
      24 CAGCTTCTCACCACTGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS61584

seq_documentation_block:
ID   AAS61584 standard; cDNA; 285 BP.
XX
AC   AAS61584;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Lung small cell carcinoma antigen, cDNA #125.
XX
KW   Human; cytostatic; antitumour; lung small cell cancer antigen;
KW   tumour; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200177168-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-US11859.
XX
PR   11-APR-2000; 2000US-196780P.
PR   21-JUN-2000; 2000US-213361P.
PR   01-SEP-2000; 2000US-229763P.
PR   05-SEP-2000; 2000US-230629P.
PR   14-SEP-2000; 2000US-232565P.
PR   19-DEC-2000; 2000US-257037P.
PR   08-JAN-2001; 2001US-260796P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Lodes MJ,  Wang T,  Mohamath R,  Indirias CY;
XX
DR   WPI; 2002-010896/01.
XX
PT   Lung tumour polynucleotide and polypeptides useful in therapy and
PT   diagnosis of cancer especially lung cancer  -
XX
PS   Claim 1; Page 166; 295pp; English.
XX
CC   The invention relates to novel isolated lung small cell cancer antigen
CC   polynucleotides (I) and polypeptides (II) used in a method of detecting
CC   cancer in a patient. The method is optionally performed by
CC   utilising oligonucleotides (III), where the biological sample
CC   from the patient is contacted with (III), detecting the amount of
CC   polynucleotide hybridised to (III) in the sample and comparing the
CC   amount of polynucleotide to a predetermined cut-off value and thereby
CC   determining cancer in a patient. (I), (II) or antigen-presenting cells
CC   expressing (II) is useful for stimulating and/or expanding T cells
CC   specific for a tumour protein. The method comprises contacting T cells
CC   with one of the components under conditions to permit the stimulation
CC   and/or expansion of the cells. A composition comprising (I) is useful for
CC   stimulating an immune response in a patient and for inhibiting the
CC   development of a cancer especially lung cancer in a patient. An
CC   isolated T cell population is useful for removing tumour cells from the
CC   biological sample and for inhibiting the development of cancer in a
CC   patient. AAS61460-AAS61874 represent novel human lung small cell
CC   cancer antigen coding sequences of the invention.
XX
SQ   Sequence 285 BP; 81 A; 59 C; 55 G; 77 T; 13 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAS61584   ..

 Align seg 1/1  to: AAS61584  from: 1  to: 285

     127 IleValProThrLeuHisLeu 133
         ||||||||||||||||||||
      98 ATAGTACCAACCCTTCATCTC 118

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:ABA11586

seq_documentation_block:
```

ID ABA11586 standard; cDNA; 295 BP.
XX
AC ABA11586;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 593.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

```
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
DR   P-PSDB; ABB15260.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Claim 1; SEQ ID NO 593; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 295 BP; 94 A; 31 C; 54 G; 114 T; 2 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA11586/rev   ..

 Align seg 1/1  to reverse of: ABA11586  from: 1  to: 295

      34 LysThrProSerValLeuIle 40
         ||||||||||||||||||||
      31 AAAACACCGTCTGTATTAATT 11

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV37035

seq_documentation_block:
ID   AAV37035 standard; DNA; 312 BP.
XX
AC   AAV37035;
XX
DT   04-SEP-1998  (first entry)
XX
DE   Streptococcus salivarius sequence used to design primers and probes.
XX
KW   Detection; bacterial antibiotic resistance gene; bacteria;
KW   fungal species; identification; ds.
XX
OS   Streptococcus salivarius.
XX
PN   WO9820157-A2.
XX
PD   14-MAY-1998.
XX
PF   04-NOV-1997;   97WO-CA00829.
XX
PR   04-NOV-1996;   96US-0743637.
XX
PA   (IDII-) IDI INFECTIO DIAGNOSTIC INC.
XX
PI   Bergeron MG,  Ouellette M,  Picard FJ,  Roy PH;
XX
DR   WPI; 1998-286967/25.
XX
PT   Use of oligo:nucleotide primers and probes - for detection,
PT   identification and quantification of bacteria, fungi and bacterial
PT   antibiotic resistance gene(s)
XX
PS   Claim 10; Page 87; 167pp; English.
XX
CC   The present sequence was used to design primers and probes which
CC   are used in the course of the invention. The specification describes the
CC   use of probes and/or amplification primers which are specific, ubiquitous
CC   and sensitive for determining the presence and amount of nucleic acids
CC   from a bacterial antibiotic resistance gene and specific bacterial and
CC   fungal species in any sample suspected of containing the bacterial or
CC   fungal nucleic acids, where each of the nucleic acid or variant or part
CC   comprises a selected target region hybridisable with the probes or
CC   primers. The method of use comprises contacting the sample with the
CC   probes or primers and detecting the presence of hybridised probes or
CC   amplified products as an indication of the presence of the specific
CC   bacterial or fungal species and bacterial antibiotic resistance genes.
CC   The methods and products can be used to detect and identify the bacterial
CC   and fungal species and genera and determine the bacterial resistance to
CC   antibiotics.
XX
SQ   Sequence 312 BP; 81 A; 54 C; 80 G; 97 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAV37035   ..

 Align seg 1/1  to: AAV37035  from: 1  to: 312

      58 SerHisValGlyLeuGlnAla 64
         ||||||||||||||||||||
     181 AGTCATGTAGGACTTCAAGCG 201

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH01713

seq_documentation_block:
ID   AAH01713 standard; DNA; 312 BP.
XX
AC   AAH01713;
XX
DT   24-JUL-2001  (first entry)
XX
DE   Streptococcus salivarius nucleotide sequence SEQ ID NO:1706.
XX
KW   Species specific; genus specific; family specific; probe; detection;
KW   identification; algal; archaeal; bacterial; fungal; parasitical;
```

```
KW   microorganism; diagnosis; translation elongation factor Tu; toxin;
KW   translation elongation factor G; RecA recombinase; resistance;
KW   catalytic subunit of proton-translocating ATPase; antimicrobial;
KW   vaccine; primer; ds.
XX
OS   Streptococcus salivarius.
XX
PN   WO200123604-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-CA01150.
XX
PR   28-SEP-1999;   99CA-2283458.
PR   19-MAY-2000; 2000CA-2307010.
XX
PA   (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX
PI   Bergeron MG,  Boissinot M,  Huletsky A,  Menard C,  Ouellette M;
PI   Picard FJ,  Roy PH;
XX
DR   WPI; 2001-245006/25.
XX
PT   Nucleic acid sequences are used to generate universal probes and
PT   primers which can be used to identify and detect the presence of algal,
PT   archaeal, bacterial, fungal and parasitical species in a test sample -
XX
PS   Disclosure; Page 1288; 1580pp; English.
XX
CC   The present invention describes a method for generating a repertory of
CC   nucleic acids of tuf, fus, atpD and/or recA genes from which probes
CC   and/or primers are derived. The method comprises amplifying the nucleic
CC   acids of determined algal, archaeal, bacterial, fungal and parasitical
CC   species with a combination of defined primer pairs. The method can be
CC   used for producing probes and/or primers for detecting one or more
CC   related microorganisms e.g. algae, archaea, bacteria, fungi and
CC   parasites, for universal detection and for specific and ubiquitous
CC   detection and identification of an algal, archaeal, bacterial, fungal
CC   and parasitical species, genus, family and group. A nucleic acid (I)
CC   obtained using the method of the invention can be used for the universal
CC   detection of any bacterium, fungus or parasite in a sample and for the
CC   detection of at least one antimicrobial agent resistance gene or at
CC   least one toxin gene. hexA nucleic acids are used for the specific and
CC   ubiquitous detection and for identification of Streptococcus pneumoniae.
CC   (I) can be used to design a therapeutic agent which is effective against
CC   microorganisms. Microbial species or genus or family or phylum or group
CC   which can be detected include Abiotrophia adiacens, Bordetella sp.,
CC   Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
CC   Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
CC   Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
CC   provides faster results than substrate specificity tests as results can
CC   be determined in an hour and improved accuracy is also achieved.
CC   AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC   which are given in the exemplification of the present invention.
XX
SQ   Sequence 312 BP; 81 A; 54 C; 80 G; 97 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAH01713    ..

 Align seg 1/1  to: AAH01713  from: 1  to: 312

      58 SerHisValGlyLeuGlnAla 64
         ||||||||||||||||||||
     181 AGTCATGTAGGACTTCAAGCG 201

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:AAT16270

seq_documentation_block:
ID   AAT16270 standard; DNA; 318 BP.
XX
AC   AAT16270;
XX
DT   01-AUG-1996  (first entry)
XX
DE   0.3 kb recA fragment from pGEM-T.
XX
KW   autonomous replication; inverted repeat; insertion sequence;
KW   open reading frame; plasmid; isolation; amino acid biosynthesis;
KW   marker gene; ds.
XX
OS   Brevibacterium flavus strain MJ-233.
XX
PN   JP07327680-A.
XX
PD   19-DEC-1995.
XX
PF   07-JUN-1994;   94JP-0124852.
XX
PR   07-JUN-1994;   94JP-0124852.
XX
PA   (MITU ) MITSUBISHI CHEM CORP.
XX
DR   WPI; 1996-072338/08.
DR   P-PSDB; R9251.
XX
PT   Plasmid which cannot be replicated autonomously in Coryneform
PT   bacteria - useful for isolating genes involved in amino acid
PT   biosynthesis
XX
PS   Disclosure; Page 15; 16pp; Japanese.
XX
CC   A plasmid which cannot be replicated autonomously in a Coryneform
CC   bacterium carries a DNA region having at least one inverted repeat
CC   (IR) present upstream or downstream of an open reading frame within
CC   an inserted sequence. The IRs are derived from the 5' upstream and
CC   3' downstream regions of a marker gene derived from a Coryneform
CC   bacterium. The plasmid is used in a method for obtaining a mutant in
CC   which a Coryneform bacterium is transformed with the plasmid and the
CC   strain expressing the marker gene is separated by using the marker gene
CC   of the transformant as the index.  Genes relating to amino acid
CC   biosynthesis can be easily isolated by using the mutant.  The present
CC   sequence is a 0.3 kb PCR fragment contg. the recA gene from
CC   Brevibacterium flavus strain MJ-233. A recA gene-deleted mutant strain
CC   was prepd. for use in the method of the invention.
XX
SQ   Sequence 318 BP; 72 A; 82 C; 91 G; 73 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAT16270    ..

 Align seg 1/1  to: AAT16270  from: 1  to: 318

      58 SerHisValGlyLeuGlnAla 64
         ||||||||||||||||||||
     187 AGCCACGTTGGTCTTCAGGCC 207

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH01012

seq_documentation_block:
ID   AAH01012 standard; DNA; 332 BP.
XX
AC   AAH01012;
XX
```

```
DT   24-JUL-2001  (first entry)
XX
DE   Bacillus cereus nucleotide sequence SEQ ID NO:1003.
XX
KW   Species specific; genus specific; family specific; probe; detection;
KW   identification; algal; archaeal; bacterial; fungal; parasitical;
KW   microorganism; diagnosis; translation elongation factor Tu; toxin;
KW   translation elongation factor G; RecA recombinase; resistance;
KW   catalytic subunit of proton-translocating ATPase; antimicrobial;
KW   vaccine; primer; ds.
XX
OS   Bacillus cereus.
XX
PN   WO200123604-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-CA01150.
XX
PR   28-SEP-1999;   99CA-2283458.
PR   19-MAY-2000; 2000CA-2307010.
XX
PA   (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX
PI   Bergeron MG,  Boissinot M,  Huletsky A,  Menard C,  Ouellette M;
PI   Picard FJ,  Roy PH;
XX
DR   WPI; 2001-245006/25.
XX
PT   Nucleic acid sequences are used to generate universal probes and
PT   primers which can be used to identify and detect the presence of algal,
PT   archaeal, bacterial, fungal and parasitical species in a test sample -
XX
PS   Claim 26; Page 960; 1580pp; English.
XX
CC   The present invention describes a method for generating a repertory of
CC   nucleic acids of tuf, fus, atpD and/or recA genes from which probes
CC   and/or primers are derived. The method comprises amplifying the nucleic
CC   acids of determined algal, archaeal, bacterial, fungal and parasitical
CC   species with a combination of defined primer pairs. The method can be
CC   used for producing probes and/or primers for detecting one or more
CC   related microorganisms e.g. algae, archaea, bacteria, fungi and
CC   parasites, for universal detection and for specific and ubiquitous
CC   detection and identification of an algal, archaeal, bacterial, fungal
CC   and parasitical species, genus, family and group. A nucleic acid (I)
CC   obtained using the method of the invention can be used for the universal
CC   detection of any bacterium, fungus or parasite in a sample and for the
CC   detection of at least one antimicrobial agent resistance gene or at
CC   least one toxin gene. hexA nucleic acids are used for the specific and
CC   ubiquitous detection and for identification of Streptococcus pneumoniae.
CC   (I) can be used to design a therapeutic agent which is effective against
CC   microorganisms. Microbial species or genus or family or phylum or group
CC   which can be detected include Abiotrophia adiacens, Bordetella sp.,
CC   Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
CC   Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
CC   Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
CC   provides faster results than substrate specificity tests as results can
CC   be determined in an hour and improved accuracy is also achieved.
CC   AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC   which are given in the exemplification of the present invention.
XX
SQ   Sequence 332 BP; 115 A; 60 C; 74 G; 83 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAH01012   ..

 Align seg 1/1  to: AAH01012  from: 1  to: 332
```

```
      58 SerHisValGlyLeuGlnAla 64
         ||||||||||||||||||||
     191 TCACACGTAGGTTTACAAGCA 211

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH01298

seq_documentation_block:
ID   AAH01298 standard; DNA; 336 BP.
XX
AC   AAH01298;
XX
DT   24-JUL-2001  (first entry)
XX
DE   Bacillus thuringiensis nucleotide sequence SEQ ID NO:1289.
XX
KW   Species specific; genus specific; family specific; probe; detection;
KW   identification; algal; archaeal; bacterial; fungal; parasitical;
KW   microorganism; diagnosis; translation elongation factor Tu; toxin;
KW   translation elongation factor G; RecA recombinase; resistance;
KW   catalytic subunit of proton-translocating ATPase; antimicrobial;
KW   vaccine; primer; ds.
XX
OS   Bacillus thuringiensis.
XX
PN   WO200123604-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-CA01150.
XX
PR   28-SEP-1999;   99CA-2283458.
PR   19-MAY-2000; 2000CA-2307010.
XX
PA   (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX
PI   Bergeron MG,  Boissinot M,  Huletsky A,  Menard C,  Ouellette M;
PI   Picard FJ,  Roy PH;
XX
DR   WPI; 2001-245006/25.
XX
PT   Nucleic acid sequences are used to generate universal probes and
PT   primers which can be used to identify and detect the presence of algal,
PT   archaeal, bacterial, fungal and parasitical species in a test sample -
XX
PS   Claim 26; Page 1105; 1580pp; English.
XX
CC   The present invention describes a method for generating a repertory of
CC   nucleic acids of tuf, fus, atpD and/or recA genes from which probes
CC   and/or primers are derived. The method comprises amplifying the nucleic
CC   acids of determined algal, archaeal, bacterial, fungal and parasitical
CC   species with a combination of defined primer pairs. The method can be
CC   used for producing probes and/or primers for detecting one or more
CC   related microorganisms e.g. algae, archaea, bacteria, fungi and
CC   parasites, for universal detection and for specific and ubiquitous
CC   detection and identification of an algal, archaeal, bacterial, fungal
CC   and parasitical species, genus, family and group. A nucleic acid (I)
CC   obtained using the method of the invention can be used for the universal
CC   detection of any bacterium, fungus or parasite in a sample and for the
CC   detection of at least one antimicrobial agent resistance gene or at
CC   least one toxin gene. hexA nucleic acids are used for the specific and
CC   ubiquitous detection and for identification of Streptococcus pneumoniae.
CC   (I) can be used to design a therapeutic agent which is effective against
CC   microorganisms. Microbial species or genus or family or phylum or group
CC   which can be detected include Abiotrophia adiacens, Bordetella sp.,
CC   Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
CC   Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
CC   Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
CC   provides faster results than substrate specificity tests as results can
CC   be determined in an hour and improved accuracy is also achieved.
CC   AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC   which are given in the exemplification of the present invention.
XX
```

SQ Sequence 336 BP; 111 A; 61 C; 75 G; 85 T; 4 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAH01298 ..

Align seg 1/1 to: AAH01298 from: 1 to: 336

```
     58 SerHisValGlyLeuGlnAla 64
        ||||||||||||||||||||
    191 TCACACGTAGGTTTACAAGCA 211
```